THE UNIVERSITY OF MICHIGAN
MAR 1 1961
BUSINESS ADMINISTRATION LIBRARY

# Journal of Advertising Research

Vol. 1, No. 3 MARCH 1961

**Sales and Advertising of Lettuce** FRANK MEISSNER 1

**How to Estimate Unduplicated Audiences** J.-M. AGOSTINI 11

**A Switch-Over Experimental Design to Measure Advertising Effect** R. J. JESSEN 15

**The Influence of Newspaper Tune-In Advertising on the Size of a TV Show's Audience** JAMES C. BECKNELL, JR. 23

**FEDERAL STATISTICS IN ADVERTISING** 27

**ON METHODS** 28

**RESEARCH IN REVIEW** 34

**PUBLICATIONS RECEIVED** 39

**EDITORIAL** 40

Published by the Advertising Research Foundation

The
# Journal of Advertising Research

is published by the Advertising Research Foundation and sent free to its members. Single copies are available to eligible non-members of ARF at $10, and to ineligible non-members at $2.50.

solicits original papers. Reports of findings are favored over theoretical discussion. Manuscripts should be submitted in triplicate, double-spaced, with references, tables and figures on separate pages. Authors receive 50 free offprints. Letters of comment and criticism are also invited.

is intended for practitioners and users of advertising research. Limits on readers' time and journal space require that papers be as short as clarity permits.

is an open forum. Publication in it implies no endorsement of the writer's purpose, methods or views by the Advertising Research Foundation, its board of directors, or any of its committees.

is edited by the ARF Technical Staff:

Charles K. Ramond ..... Technical Director
Ingrid C. Kildegaard ..... Research Statistician
Gwyn Collins ..... Research Mathematician
Thornton C. Lockwood ..... Research Associate
George M. Shirey ..... Research Associate
Mildred Cantrell ..... Librarian
Joey Harris ..... Secretary

© 1961 by the Advertising Research Foundation, Inc.
3 East 54th Street, New York 22, N. Y.

# Sales and Advertising of Lettuce[1]

FRANK MEISSNER
San Jose State College

**For several years California growers promoted nationally a new brand of lettuce, using as many as eight media. Ratings of two of these media, along with local retail price and temperature, were observed to be closely correlated with per capita sales in 22 cities.**

*Two businessmen were talking.*

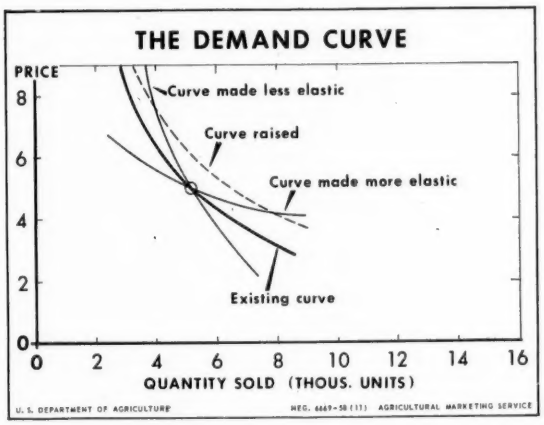
*"Is your advertising getting results?" inquired the first.*

*"It sure is!" moaned the second. "Last week we advertised for a night watchman and the next night we were robbed."*

THE LONG-RANGE PURPOSE of advertising and sales promotion is to persuade consumers to buy more of a commodity at the same price, or to pay higher prices for the same quantity, or to do a little of each. In the economist's language, advertising effect can be gauged in terms of shifts either in the position or the slope of the demand curve for the product (see Figure 1). This theoretical framework has been reviewed elsewhere (Fox, 1959; Hoos, 1959; and Waugh, 1959).

Conceptually, the measurement of advertising impact appears to be a straightforward task. However, in reality it is quite a difficult job, because advertising is only one of many "demand shifters."

FIGURE 1



**FRANK MEISSNER is assistant professor of economics at San Jose State College. A native of Czechoslovakia, he received a B.Sc. from the Royal Agricultural College in Copenhagen, an M.S. from Iowa State University and a Ph.D. in marketing from Cornell University in 1957. After three and a half years at Crown Zellerbach in charge of marketing research for the distributor sales division, he worked for a year as an industrial economist at the Stanford Research Institute, where he is still employed as a consultant. He has published widely here and abroad, and serves as an assistant editor of the Marketing Articles in Review section of the *Journal of Marketing*. He is writing a book on measuring the effect of advertising.**

[1] Drafts of this paper were read by Thomas A. Mainwaring, senior economist, Stanford Research Institute; Walter J. Englund, marketing economist, Bureau of Markets, California Department of Agriculture; and Karl V. Christierson, vice president, Major Farms, Salinas, California. Their comments were greatly appreciated. Sole responsibility for the contents of the paper remains the present author's.

Changes in income, taste, habits, prices and quantities of competing products are some other important factors that affect demand.

This unsatisfactory state of the art is clearly reflected in the views expressed by the majority of respondents in a recent survey of executives in 152 manufacturing firms made by the National Industrial Conference Board (Mathes and Thompson, 1960). Their gloomy attitudes were confirmed by participants in a recent workshop on "measuring advertising effectiveness" held by the Northern California Chapter of the American Marketing Association. "The consensus of the group was: 'You can't do it.' There are too many factors involved to [permit] realistic research. . . ." (*Newsletter*, Northern California Chapter of AMA, 1960, p. 1).

But it would appear that the great number of variables is by no means a fatal complication. After all, even a reasonably straightforward trend analysis incorporates in the model, almost as a matter of routine, several variables such as mentioned above. The parameters of demand models are being estimated by means of such simple and well known methods as least squares regression and correlation, or perhaps by the more sophisticated tools of econometric analysis using systems of simultaneous equations.

### *Applications*

In recent years several attempts have been made to apply these analytical tools to the measurement of advertising impact on sales of food and farm products. The slow going seems due far more to inadequate data on advertising and other inputs rather than to the inherent methodological difficulty of the problem.

For example, one of the most mathematically elegant analyses (Baseman, 1955) found that tobacco consumption was more responsive to prices than to advertising. A least squares demand function showed that during the 20 years 1929-1950, a one per cent increase in industry-wide advertising expenditure led to an average one-twentieth of a per cent increase in tobacco consumption, while a one per cent decrease in prices boosted consumption by one-fifth. Also estimated were demand and marginal utility elasticities by the method of simultaneous equations. No attempt was made to show how manipulations of an individual firm's advertising budget would have influenced the demand for the firm's own brand or brands.

## PURPOSE

This paper has two objectives: first to review an advertising evaluation study of lettuce which used the multiple regression method of analysis; and second to point out the direction which future research might take.

Lettuce is used for illustrative purposes primarily because past work was readily available (Mainwaring et al., 1955). Furthermore, lettuce is a relatively "easy" commodity for which impact of producer-sponsored advertising can be isolated.

### *Production*

The great bulk of commercial production is concentrated in rather well defined areas. Thus in 1954-58 the national production of lettuce averaged about 1,630,000 tons annually, of which about 990,000 tons, or almost two-thirds, were produced in California.

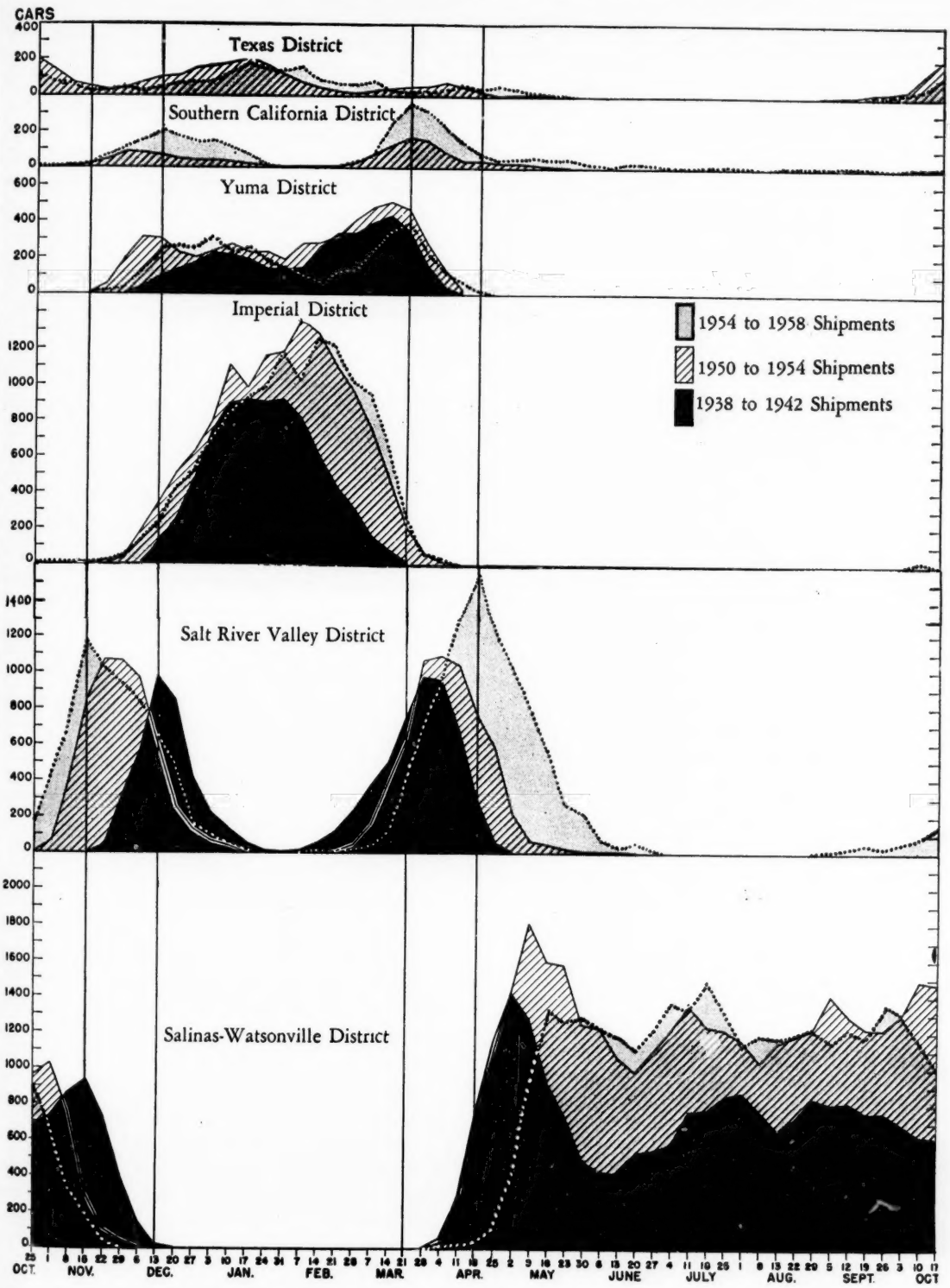
California lettuce is marketed during two seasons from two major growing districts. The Imperial Valley marketing season is from November through March, the Salinas season from April through November. The Salinas district in the Central Coast area ships the largest volume—about two-thirds of all California lettuce or nearly one-half of the total rail shipments of lettuce in the United States. From April through November shipments from the Salinas district dominate the U. S. lettuce trade. However, in recent years shipments from the Salt River Valley area in Arizona have been "cutting into" Salinas lettuce shipments particularly during November and April (see Figure 2).

### *Consumption*

In the last four decades lettuce has developed into one of the most important vegetables. In 1954-1958, per capita consumption was approximately 19 pounds, or almost four times higher than in the immediate post-War I years. Lettuce's share of the total value of vegetables produced for the fresh market increased from about ten per cent in 1920 to roughly 17 per cent in the most recent five years. In 1958, the farm value of lettuce marketed in the United States amounted to about $130 million.

This remarkable growth in lettuce consumption is due to a combination of several favorable factors: changing dietary habits resulting in emphasis on non-fats; increased levels of living which allow consumers to buy foods once considered luxuries; fewer hours of hard manual labor which reduced

FIGURE 2: AVERAGE WEEKLY SHIPMENTS OF LETTUCE (HOPKIN, 1960)



need for high energy foods; increased population; new lettuce varieties; and technological innovations which have made an assembly line operation out of lettuce growing: field packing, vacuum cooling and faster shipping.

The per capita consumption of lettuce will grow further within the next decades for two reasons. First, many low-income consumers still eat little lettuce; as per capita income continues to grow and become more evenly distributed, more people will eat more lettuce. Second, institutional outlets (hotels, restaurants, schools, hospitals) are becoming increasingly important. On the average, relatively more lettuce is eaten in these places than in meals eaten at home.

Combinations of these favorable factors will probably boost the per capita consumption of lettuce to 22-25 pounds by 1975. A successful advertising and sales promotion campaign would tend to accelerate this upward trend, and might therefore show a "plus" effect. This type of evidence is more convincing than if advertising tends to slow down a declining trend.

### *The C-7 Campaign*[1]

In 1952 a group of lettuce shippers of Salinas, California formed Lettuce, Inc., a non-profit corporation, in order to increase the consumption of lettuce by nationally advertising and promoting their C-7 brand (C for California, and 7 for seven days a week). In attempting to achieve this goal, part of the program was directed toward influencing members of the produce trade to merchandise lettuce more effectively by: considering lettuce as the most basic commodity in their fresh produce line; using lettuce for promotional purposes; prepackaging lettuce; and participating in the C-7 program by handling C 7 lettuce, using C-7 advertising and point-of-purchase materials, and using C-7 trademarked cellophane wrappers for prepackaging C-7 lettuce.

From 1952 to 1957 the C-7 campaign consisted partly of spot or network radio advertising. Television time on the "Home" program was bought in 1955. In late 1955 and in 1956 the "Mickey Mouse Club" was used. Promotion efforts were carried out by fieldmen from Lettuce, Inc., who solicited support for C-7 at wholesale as well as retail levels in major population centers of the United States; their work was supplemented by direct mailings. Point-of-sale promotion materials were made available, together with prepackaging wrappers with the C-7 label imprint. Newspaper advertising mats were also supplied to retailers.

The sustained campaign was handled by John Cohan Advertising in Salinas. The expenditures for the campaign increased from about $150,000 in 1952 to about $420,000 in 1956.

In 1956 the Board of Directors of Lettuce, Inc. commissioned Stanford Research Institute to evaluate the effect of their promotional campaign upon lettuce consumption. A proviso was issued to "let the chips fall where they may."

### METHOD

The impact of the campaign was evaluated by two methods.

First, a series of interview surveys was made with lettuce brokers, wholesalers, retailers and consumers in two cities which differed primarily in the amount of C-7 advertising and promotion received per capita: Grand Rapids, Michigan, which received relatively much attention, and Butte, Montana, which received little attention. F.o.b. buyers of lettuce in the Salinas area of California were also interviewed.

Second, a multiple regression analysis was carried out for 22 cities: Atlanta, Baltimore, Boston, Chicago, Cincinnati, Cleveland, Dallas, Denver, Detroit, Fort Worth, Jackson (Mississippi), Kansas City (Missouri), Los Angeles, New Orleans, New York, Oakland, Philadelphia, Pittsburgh, St. Louis, San Francisco, Seattle and Washington, D. C. This was done in an attempt to explain the city-to-city variation in per capita consumption of lettuce (dependent variable) in terms of various demand shifters (independent variables).

In other words, it was assumed that fluctuations in per capita consumption of lettuce in a certain marketing area can be explained by the following functional relationship:

$$L = f(I,T,P,A\ldots)$$

This symbolic expression is shorthand for the following idea: per capita consumption of lettuce (L) in a certain marketing area depends primarily on (is a function of) the area's per capita income (I), temperature (T), retail price of lettuce (P) and the intensity of the advertising effort (A). Other variables that cannot readily be quantified are indicated by the dots.

[1]This section of the paper was discussed at the 120th annual meeting of the American Statistical Association at Stanford University August 26, 1960 by H. S. Houthakker and F. Meissner.

TABLE 1

FACTORS ASSOCIATED WITH VARIATION IN PER CAPITA CONSUMPTION OF HEAD LETTUCE, 22 SELECTED U.S. CITIES, 1950-55 (SOURCE: STANFORD RESEARCH INSTITUTE)

| *Variables* | *Units* | *Mean* | *Coefficients* | |
|---|---|---|---|---|
| | | | *Partial Correlation* | *Regression* |
| *Independent:* | | | | |
| 1. Income | $10,000 per capita per year | .1493 | .129 | .1789 |
| 2. Price | Dollars per head of lettuce | .1345 | —.595* | —1.1159 |
| 3. Temperature | 100° Fahrenheit | .7962 | —.228* | —.1321 |
| 4. Radio spot | 100's of broadcasts | .0161 | —.057 | —.0512 |
| 5. Radio network | 100's of broadcasts | .1183 | —.154 | —.0591 |
| 6. Television | 100's of broadcasts | .0212 | .120 | .1415 |
| 7. Fieldmen | Ratings of effectiveness | .0346 | —.201* | —.2905 |
| 8. Wholesalers | No. per 100,000 persons | .0380 | .091 | .0428 |
| 9. Retailers | No. per 1,000 persons | .0276 | .120 | .1838 |
| 10. Direct mail | No. per 1,000 persons | .0048 | .152 | 1.2109 |
| 11. Newspapers | Ratings of cooperation | .0203 | .237* | .4578 |
| *Dependent:* | | | | |
| Unloads of lettuce | Carlots per 100 persons | .1107 | | |

* Significant beyond the .10 level of confidence.

1. *Income:* Effective Buying Income from "Survey of Buying Power," *Sales Management* magazine. The original data were adjusted to a 1947-49 constant dollar basis by the annual average consumer price index for the respective city, or for the nearest comparable city.
2. *Price:* Monthly averages from *Retail Food Prices by Cities,* Bureau of Labor Statistics, U.S. Department of Commerce. *The Consumer Price Index,* Series B-10 of the U.S. Department of Commerce provides comparable figures. The 1947-49 deflated monthly prices were weighted by monthly unloads to provide annual averages.
3. *Temperature:* Averages of daily maxima, Local Climatological Summary, U.S. Weather Bureau.
4-5. *Radio:* Number (in hundreds) of spot and national network broadcasts.
6. *Television:* Number (in hundreds) of network broadcasts of the "Home" and "Mickey Mouse" programs.
7. *Fieldmen:* Client rating: .10 (maximum), .07 (moderate), .03 (very little) and 0 (no contact).
8. *Wholesalers:* Number of service wholesalers who actively supported the C-7 program.
9. *Retailers:* Number of retailers who used promotions in their stores and ran newspaper ads featuring C-7 lettuce.
10. *Direct Mail:* Number of names on the C-7 mailing list. Approximately six bulletins were sent annually, announcing special events or asking for support of the C-7 program.
11. *Newspapers:* Client rating of cooperation received: .10 (maximum), .07 (moderate), .03 (very little) and 0 (none).

The population variable was eliminated by expressing most data on a per capita basis. Radio and TV advertising broadcasts were retained in original terms, because the media coverage was proportional to population. Data on monetary expenditures on individual media were not available. One single expression of the sales promotion and advertising effort could not be developed. Time was excluded, because the partial correlation coefficient corresponding to the time variable was only .052, or insignificantly small. All calculations were carried out with a standard program on the IBM 650 electronic computer.

## RESULTS

Results of the regression analysis are shown in Table 1. Shown for all variables are: units of measurement, means of observations, partial correlation coefficients and regression coefficients.

The partial correlation coefficient indicates the "closeness of association" between an individual explanatory variable and lettuce consumption, after adjustments have been made for effects of other explanatory variables. The regression coefficient indicates the association between a unit change in the explanatory variable and lettuce consumption. Minus signs in front of the coefficients indicate a negative relation between the independent and dependent variables. For instance, everything else being equal, an increase of one per cent in the retail price of lettuce led to a decrease of about one per cent in lettuce unloads.

It appears that in 1950-55 only four variables—price, temperature, fieldmen and newspaper advertising—had a significant influence on lettuce consumption. Of the remaining six advertising variables, it cannot be said with confidence that they had a net effect on lettuce consumption. This does not mean that the C-7 program as a whole had no effect. On the contrary: the multiple correlation coefficient between per capita lettuce consumption and all 11 explanatory variables was .66, but merely .60 if only price, income and weather were included. This means that the chances are better than 1:19 that the eight advertising variables had a significant joint effect.

Specifically, it appears that at prevailing prices the C-7 program tended to increase demand for head lettuce by an annual equivalent of about 8.8 carlots per 100,000 population. (The sum of the products of individual regression coefficients and their means came out to about seven carlots per 100,000 persons per year. This estimate was increased by 27 per cent to allow for the fact that 22 out of the 104 individual observations referred to years 1950 and 1951, when the C-7 program was not in effect.)

Instead of a single equation approach, a simultaneous equations technique might have been more appropriate. But in view of the relative crudeness of the original data inputs, it did not appear that the results would have been sufficiently more reliable to warrant the extra time and effort required by this rather sophisticated analytical tool.

The C-7 campaign was the only national advertising activity for lettuce carried out at that time. Thus no adjustment for competitive promotions by lettuce producers in other areas was necessary.

The results of the statistical analysis were consistent with the findings of the interview surveys, which showed that:

1. Among consumers in areas reached by the campaign, an awareness of the C-7 campaign and demand for C-7 lettuce was created.

2. The campaign influenced wholesalers and retailers to increase their selling effort on lettuce in general, and on C-7 lettuce in particular.

A similar interview survey has also been done for asparagus (Mason, 1959). It showed that current advertising and promotional efforts have stimulated sales, yet trade members have no real desire to push asparagus without industry aid. While proceeding toward these conclusions, the study, commissioned by the California Asparagus Growers, unearthed extensive information on methods of promotion, merchandising and handling, and on the competitive position of California asparagus. It also provides model interview techniques for generating comprehensive marketing and distribution data.

## DISCUSSION

An examination of the above analysis points toward interesting areas of future explorations.

### *Other Variables*

The independent variables used in the multiple regression analysis explained only about half of the total variation in the per capita consumption of lettuce: 38 per cent by price, income and weather, and an additional 12 per cent by the eight advertising variables (see Table 2). It is therefore evident that important factors which affect consumption of lettuce were left out of the analysis. An attempt should be made to include additional variables that would narrow down the unexplained or residual part of the variation.

TABLE 2

ANALYSIS OF VARIANCE IN PER CAPITA CONSUMPTION OF LETTUCE

| *Source of Variance* | *Per cent of Total Variation*[1] | *Degrees of Freedom* | *Mean Squares* | *F Ratio* |
|---|---|---|---|---|
| Non-advertising (variables 1-3 in Table 1) | 38.0 | 100 | — | |
| Advertising (variables 4-11 in Table 1) | 11.8 | 8 | .0148 | 2.69 |
| Total explained variation | 49.8 | 92 | — | |
| Total unexplained variation | 50.2 | 92 | .0055 | |

Source: Stanford Research Institute

[1] The partial correlation coefficients, which in Table 1 are adjusted for degrees of freedom, were converted back to unadjusted form, and then squared.

### *The Returns Curve*

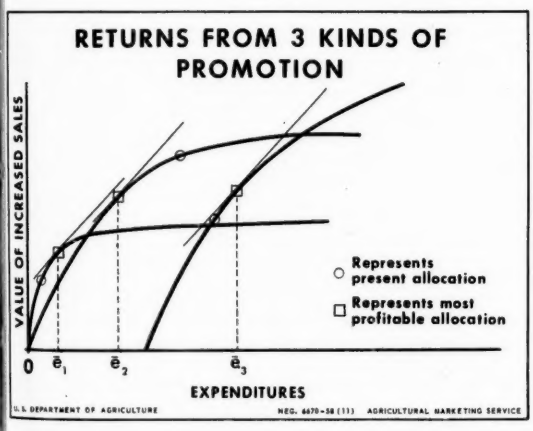
Lettuce growers and shippers—often the same person or firm—advertise and promote so as to increase net revenue per crate f.o.b. shipping point and to broaden the market for lettuce. (An expression for f.o.b. shipping point prices would have

been helpful in supplementing the retail prices used in the original Stanford Research Institute study.) In other words, they want to know either how best to allocate a given advertising and promotion expenditure among geographic areas and media so as to *maximize sales,* or how to *minimize advertising and promotion effort* while achieving a given volume of sales.

For this purpose one needs to estimate the shape and place of the "returns curve" for each area, which would show "the expected net value of increased sales due to promotion." This principle was suggested more than two decades ago by Dr. Harry R. Wellman (1939) now vice president of the University of California. Wellman originally used the term "sales curve" to mean the relationship between the expected volume of sales and varying levels of promotional expenditures in a marketing area. Waugh (1959) prefers "returns curve."

The returns curve principle is illustrated in Figure 3. Conceptually the three curves might represent the expected increases in the net returns for lettuce in, say, three of the 22 marketing areas for which consumption data are available. All curves tend to flatten off; this is to indicate that there probably would be decreasing returns to promotion.

FIGURE 3




Here is the functioning of this model as explained by Waugh (1959, pp. 211-212):

> Suppose that the initial allocation of promotional funds is that shown by the circles. At these three points, the middle curve is flatter (less steep) than either of the others. The slopes of the curves measure marginal returns. In other words, the initial allocation of promotional funds, the marginal returns from market 2, are less than those from either market 1 or market 3. Suppose that a dollar of promotional funds were transferred from market 2 to market 1. The loss in net returns from market 2 would be more than offset by the gain from market 1. The same is true of a transfer of a dollar of promotional funds from market 2 to market 3. Such transfers of funds would always be profitable so long as the marginal returns from any pair of markets were unequal. Thus, the equalization of marginal net returns is a necessary condition for maximum net income.
>
> One could shift funds and observe graphically the slopes of the several curves. In time, the slopes of all would be equal—that is, where tangents to the curves were parallel. In Figure 3, such a situation is shown by the squares, corresponding to expenditures $\bar{e}_1$, $\bar{e}_2$, $\bar{e}_3$, in markets 1, 2, and 3. Assuming that the three curves accurately measure the responses to advertising in three markets, and assuming a promotional budget of $(\bar{e}_1 + \bar{e}_2 + \bar{e}_3)$ dollars, the most profitable allocation of that budget is $\bar{e}_1$, to market 1, $\bar{e}_2$, to market 2 and $\bar{e}_3$, to market 3.
>
> *No one can successfully dispute this principle. It is a geometric fact, like the Pythagorean theorem about right triangles. It is true—exactly, inescapably, eternally. The only question is how to estimate the returns curves. This must be done, in one way or another, if the problem of optimum allocation is to be solved.* (Italics added—F.M.)

Chances are that the sets of multiple regression analyses (demand curves) discussed previously will provide reasonable estimates of the returns curve for lettuce.

Series of regression analyses have already been successfully used in an attempt to explain the fluctuations in f.o.b. shipper prices for western lettuce. Shipments from different production regions in the west, as well as from competing areas, and consumer income were used as explanatory variables (McGlothlin, 1957; Brennan and Toussaint, 1958).

### *The Decay Curve*

The C-7 campaign was discontinued in 1957. This fact alone opens up new possibilities for empirical testing of the concept of the decay curve (also called distributed lags; see Friedman, 1957, and Nerlove, 1958) and its application to measuring the effect of advertising and sales promotion.

Past work in this field has primarily focused on *current* relations between advertising and sales. This type of analysis is appropriate for evaluation of short time promotions such as weekend specials in food stores. However, it seems that most advertising that has been sustained over a long period of time does not show a significant effect till some time after the campaign has started. By the same token, it has some effect for months and perhaps years after the campaign ceased. In fact, recent studies (Alderson, 1958; Vidale and Wolfe, 1957) indicate that "old ads never die; they just fade away."

All this suggests that the multiple regression analysis on lettuce should perhaps be "decomposed" into a number of equations, rather than one catch-all relationship. For instance, it appears possible to construct a series of demand curves by quarterly or semi-annual intervals for the years 1950 through 1960. This would probably provide some indication of the gradual demand-shifting effect of the C-7 campaign. Equations for the 1950-52 years could serve as a "before" or control period; 1953-57 might be used as a "during" period in which the build-up of the advertising effect could be traced; and the 1958-60 "after" period would presumably help to outline the decay curve.

### *Media Evaluation*

Marginal returns will have to be equalized not only between individual marketing areas but between advertising media as well. Many problems would be encountered, because data on advertising inputs are usually not available in sufficient detail. The problem can be minimized should advertising agencies alter slightly their cost accounting so as to retain enough detail.

The analysis shown in Table 1 has produced a number of interesting results which indicate the kind of information which media analysis could provide. For instance, although the joint effect of all eight of the promotional variables is positive, the regression coefficients of some of the variables are actually negative. For radio spot and network advertising this may mean that these media tended to discourage lettuce consumption. But since neither of the coefficients is significant, such a statement is not warranted without additional investigation.

However, for fieldmen the regression coefficient is negative and significantly so. Here again it cannot be concluded that promotion by fieldmen actually tended to decrease lettuce consumption. This is because the fieldmen worked indirectly by trying to induce wholesalers and retailers to cooperate by advertising in newspapers (using mats provided by Lettuce, Inc.) and in many other ways. Thus the regression coefficients for wholesalers, retailers and newspapers are positive. This can largely be traced back to fieldmen, because when these three variables are left out from the equation, the effect of fieldmen turns out to be positive.

In the following section we explain an experimental method by which such interactions could be systematically isolated.

### *Systematic Experiments*

A mathematical model based on historical data can of course merely help to explain the past. It might have only limited predictive value. Yet the advertiser wants to know, within reasonable limits of error, what he can expect in the future if he allocates his funds in certain ways. To provide such knowledge the mathematical models will have to be supplemented with results from controlled marketing experiments. In such experiments the size and type of advertising investments can be systematically varied, the influence of differences in location and time can largely be eliminated, and the interactions between media can be isolated.

This approach is analogous to laboratory experiments in which the scientist measures the types of changes brought about in the dependent variable by systematically controlled changes in the causative (independent) variable. It is not necessary to test the effect of one independent variable at a time; different experimental designs for analysis of variance are available, which make it possible to simultaneously determine the effect of several variables. These designs are being routinely employed in physical as well as biological sciences. In fact, within the last decade some of the complex cross-over or rotational latin square designs (Brunk, 1953; Brunk and Federer, 1953; Wood, 1960) have been successfully used in a long series of merchandising studies carried out under the auspices of the U. S. Department of Agriculture (Goldsborough, 1959; Stockman, 1959; Wells, 1959). The experiments have primarily focused on measurement of the sales impact of alternative merchandising techniques in the fields of: packaging (type, size, color), display (type, size, location), product quality and price (Dominick, 1960). Curiously enough, one of the earliest studies (Smith, 1955) dealt with an evaluation of the impact of four methods of merchandising lettuce (bulk displays with two-head price, bulk displays with one-head price, cellophane bags with one- and two-head prices, and banded with semimetal tape with one- and two-head prices) on sales in 12 retail food stores in Pittsburgh.

Since the lettuce study, which also included experiments with apples and tomatoes, work has been done on pears (Huelskamp, 1957; Smith et al., 1955), cherries, carrots, bananas (Smith et al., 1956), peanut butter, creamery butter (Huelskamp, 1957), dates (Smith and Havas, 1959), canned foods (Grubbs et al., 1960) and pork (Naumann et al., 1959).

By and large, the U. S. Department of Agriculture studies have focused on short-range effects of specific merchandising or promotional devices. In contrast, an attempt to compare systematically the effects of two different advertising themes was conducted in 1957 by the Washington State Apple Commission (Henderson, 1959). A rotational latin square experiment was used in six midwestern cities. The following three "treatments" were tested for four-week periods in each city: a general health theme, a specific apple advertising theme, and no advertising. In other words, at any particular time during the experiment all three treatments were conducted simultaneously in three pairs of the six cities, with ten of 12 retail stores in each city. Data on sales and prices of apples and competing fruits and competitive advertising effort were also obtained.

## CONCLUSIONS

*An American approaches a Frenchman near the Seine. "Do you speak English?" the American asks. "A little," the Frenchman replies. "How deep is this river?" the American asks. "I'm sorry, I don't know." "How fast is the current?" "I don't know that either." "Well, how far is it across?" "I'm sorry." "You don't know much about this river, do you?" "It has beauty, and it's soft and it moves," the Frenchman replies. "On gray days, the river, too, is gray. Yet, when the sun sets, the water looks golden. This much I know. This is important to me."*

—Europe, 1960, *Look*, August 16, 1960, p. 17.

1. It seems that the regression method holds promise as a tool for quantitative evaluation of the sales producing effect of advertising. Chances for success appear particularly good for commodities with certain attributes similar to those of lettuce: a major part of total production is confined to a few rather well defined areas; per capita consumption trend is upward; and the advertising campaign is being carried out by one group of sponsors, without simultaneous confounding effects from other promotions. Products such as California avocados, clingstone peaches, walnuts, cantaloupes, turkeys, fresh oranges and lemons, artichokes and dates would fall within this category.

2. The Stanford Research Institute evaluation was requested only after the campaign had been going on for several years. By then a number of statistics were unintentionally "let out of the barn," and could not be rounded up again in suitable form. This experience clearly indicates that for best results the process of evaluation should be incorporated into the campaign plan. This means that the data on advertising inputs (i.e., expenditures on individual media) should be broken down into geographic areas that coincide as closely as possible with the marketing areas for which sales, population, income, price and other such data are available. In most cases this type of matching of input and output data will require practically no additional bookkeeping. Yet it would increase considerably the reliability by which statisticians would be able to explain "what has happened." If the sales data were matched with qualitative measurements of, say, attitude, preference, awareness—as was done in the case of lettuce—the explanatory power of the statistical analysis could be enhanced still further.

3. Although careful historical analyses will prove extremely useful, they alone have only a limited predictive value. In order to be able to say anything about "what most probably will happen" controlled marketing experiments will be required.

4. Finally, a word of caution. A careful integration of the currently available historical quantitative and qualitative techniques with a systematic program of controlled experimentation will build up a body of general principles. These would help to bring the evaluation of advertising further into the realm of science. This does not mean that advertising will entirely cease to be an art and suddenly become a matter of mechanics. Far from it. The element of creativeness and inspiration required in generating crucial hypotheses is often a matter of unexplainable flashes of insight—or artistic inspiration, if you please.

In other words, we cannot afford to let ourselves be blinded by the promising possibilities of quantitative measurement. A balanced view of the role of advertising obviously requires inclusion of the perhaps somewhat more esoteric yet equally crucial contributions that are made by the experienced practitioners in the behavioral sciences. The psychologist, the anthropologist and the sociologist help in the process of choosing media, structuring the form and substance of the message, in building up awareness and eventually creating the motivation to buy. Without properly reflecting this often unquantifiable leg-work that precedes a sale, the statistician would be just another panacea peddler. He might well end up with little to measure, and nothing to enjoy.

## REFERENCES


ALDERSON, WROE. The Productivity of Advertising Dollars. *Cost and Profit Outlook*, Vol. 11, No. 3, February 1958.

AMERICAN MARKETING ASSOCIATION, NORTHERN CALIFORNIA CHAPTER. *Newsletter*. American Marketing Association, 1960.

BASEMAN, ROBERT L. *Application of Several Econometric Techniques to a Theory of Demand with Variable Tastes*. Unpublished Ph.D. dissertation. Ames: Iowa State University Library, 1955.

BRENNAN, N. CARSON AND W. D. TOUSSAINT. *An Economic Analysis of North Carolina Lettuce Marketing Practices*, Information Series #64. Raleigh: North Carolina Department of Agricultural Economics, February 1958, pp. 34-36.

BRUNK, MAX E. Controlled Experiments in Retail Merchandising. *Journal of Farm Economics*, December 1953, pp. 916-923.

BRUNK, MAX E. AND WALTER T. FEDERER. Experimental Design and Probability Sampling in Marketing Research. *Journal of the American Statistical Association*, Vol. 48, No. 263, September 1953, pp. 440-452.

DOMINICK, BENNETT A. Research in Retail Merchandising of Farm Products: Appraisal of Methods and Annotated Bibliography. In *Marketing Research Report #416*. Washington, D. C.: U. S. Department of Agriculture, July 1960.

FOX, KARL A. *Demand Expansion and Agricultural Adjustment*. Ames, Iowa: Center for Agricultural Adjustment, Spring 1959, pp. 123-153.

FRIEDMAN, MILTON. *A Theory of the Consumption Function*. Princeton: Princeton University Press, 1957.

GOLDSBOROUGH, GEORGE H. Experience in Designing Research Aimed at the Evaluation of Agricultural Promotion Programs. In *Promotion of Agricultural Products: Techniques, Evaluation, Theory, Policy*. Conference Proceedings of Committee on Agricultural Marketing Research of the Western Agricultural Economics Research Council, April 1959.

GRUBBS, VIOLET et al. Effectiveness of Selected Canned Food Displays in Supermarkets. In *Methodology Supplement to Agricultural Marketing Reports #371*. Washington, D. C.: U. S. Department of Agriculture, April 1960.

HENDERSON, PETER L. *Measuring Effect of Promotional Programs on Sales of Apples and Other Selected Agricultural Commodities*. A talk presented to the Washington State Apple Commission's Marketing Advisory Committee, Wenatchee, Washington, November 10, 1959.

HOOS, SIDNEY. The Advertising and Promotion of Farm Products—Some Theoretical Issues. In *Advancing Marketing Efficiency* (Lynn H. Stockman, Ed.). Chicago: American Marketing Association, 1959, pp. 189-214.

HOPKIN, JOHN A. *Commercial Vegetable Production in California*. San Francisco: Bank of America, 1960.

HUELSKAMP, HENRY J. Evaluation of Retail Merchandising Practices for Pears, Peanut Butter and Creamery Butter. In *Agricultural Marketing Report #180*. Washington, D. C.: U. S. Department of Agriculture, June 1957.

MAINWARING, THOMAS A. *An Evaluation of the Effectiveness of the Advertising and Promotion Program Conducted by Lettuce, Incorporated*, Project No. I-1800. Palo Alto, Calif.: Stanford Research Institute, 1956.

MASON, R. HAL. *California Asparagus Marketing Study*, Project No. I-2793. Palo Alto, Calif.: Stanford Research Institute, September 1959.

MATHES, SORRELL M. AND G. CLARK THOMPSON. The Art and Science of Advertising Measurement. *Business Record*, February 1960, pp. 22-25.

MCGLOTHLIN, ROBERT S. *Price Relationships in the Western Lettuce Industry*, Bulletin #287. Tucson: Arizona Agricultural Experiment Station, June 1957.

NAUMANN, H. D. et al. *A Large Scale Merchandising Experiment with Selected Pork Cuts*, Research Bulletin #711. Columbia: Missouri Agricultural Experiment Station, August 1959.

NERLOVE, MARC. *The Dynamics of Supply*. Baltimore: Johns Hopkins University Press, 1958.

SMITH, HUGH M. *Merchandising Studies in Supermarkets—Apples, Lettuce, and Tomatoes (A Preliminary Report)*. Washington, D. C.: Agricultural Marketing Service, U. S. Department of Agriculture, March 1955, pp. 5-9.

SMITH, HUGH M. AND NICK HAVAS. Displaying Dates in Packages of Different Sizes. In *Agricultural Marketing Report #349*. Washington, D. C.: U. S. Department of Agriculture, November 1959.

SMITH, HUGH M. et al. Merchandising of Selected Food Items in Grocery Stores—Canned Red Sour Cherries, Carrots, and Bananas. In *Agricultural Marketing Report #111*. Washington, D. C.: U. S. Department of Agriculture, February 1956.

SMITH, HUGH M. et al. Merchandising Winter Pears in Retail Food Stores. In *Agricultural Marketing Report #102*. Washington, D. C.: U. S. Department of Agriculture, September 1955.

STOCKMAN, LYNN H. Economic Theory and the Promotion of Agricultural Products. In *Promotion of Agricultural Products: Techniques, Evaluation, Theory, Policy*. Conference Proceedings of Committee on Agricultural Marketing Research of the Western Agricultural Economics Research Council, April 1959.

VIDALE, M. L. AND H. B. WOLFE. An Operations Research Study of Sales Response to Advertising. *Operations Research*, Vol. 5, No. 3, July 1957.

WAUGH, FREDERICK V. Needed Research on the Effectiveness of Farm Products Promotion. In *Advancing Marketing Efficiency* (Lynn H. Stockman, Ed.). Chicago: American Marketing Association, 1959, pp. 189-214.

WELLMAN, HARRY R. The Distribution of Selling Effort Among Geographic Areas. *Journal of Farm Economics*, January 1939.

WELLS, O. V. Agricultural Market Promotion—Problems, Scope, and Policy. In *Promotion of Agricultural Products: Techniques, Evaluation, Theory, Policy*. Conference Proceedings of Committee on Agricultural Marketing Research of the Western Agricultural Economics Research Council, April 1959.

WOOD RESEARCH CORPORATION. The Measurement of Advertising Effect. Philadelphia: *Wood Chips*, May 1960.

# How to Estimate Unduplicated Audiences

**J.-M. AGOSTINI**
**Elvinger Advertising Agency**

**Monsieur Agostini has developed a formula which may render unnecessary the laborious tabulation of the unduplicated audiences of combinations of three or more media vehicles. It has been successfully applied to both French and U. S. magazine data.**

WHEN SEVERAL VEHICLES (magazines, newspapers, etc.) are used in an advertising campaign, it is easy to sum their audiences. But because of audience duplication, this sum A always exceeds the audience C actually reached by the combination, and it is precisely this unduplicated audience C which must be measured in order to evaluate the coverage of the campaign.

Audience analyses sometimes give data on audience duplication for vehicles taken two by two; seldom do they give data on audience duplication for a combination of three or more. But the size of the unduplicated audience of several vehicles depends not only on the number of people simultaneously covered by two, but also on the number simultaneously covered by 3, 4 . . . . n vehicles. Analyses made in the U. S. and in France have shown that readership of three or more vehicles, particularly magazines, is very common.

Therefore, when more than two vehicles are to be used, as usually happens, the unduplicated audience is rarely known. However, the press readership survey made in France by the Centre d'Etude des Supports de Publicité (CESP) in 1957 may be considered a remarkable effort in this field. Duplications were obtained for 30 magazines taken two by two—as well as for all possible combinations of 15 of the 30 magazines! Three large books were required to present the results of the 32,767 combinations possible.

These data are obviously very useful, but refer only to half the publications covered by the survey. They do not show the unduplicated audience of any magazine combination which includes a magazine not among the 15 chosen.

The purpose of this paper is to describe a shortcut method of estimating the unduplicated audience of a higher order combination of vehicles (trio, quartet, etc.) from data on the duplication of these vehicles taken two by two.

When applied to the CESP results, the proposed method leads to a satisfactory evaluation of the unduplicated audience provided by any of the 1,073,741,793 combinations possible with the 30 magazines whose paired duplications are known.

**JEAN-MICHEL AGOSTINI is marketing consultant and a director of the Elvinger Advertising Agency in Paris. He received the degree of Ingénieur des Arts et Manufactures from the Ecole Centrale de Paris in 1942. After several years as a consultant in market research, he joined Elvinger in 1951 as director of research before assuming his present position. He is vice president of the Association Nationale des Practiciens en Etudes de Marché, member of the board of the Association pour le Developpement des Techniques d'Etudes de Marchés, chairman of the Collège Echanges d'Expériences of the Institut de Recherches et d'Etudes Publicitaires (IREP) and a member of the technical committee of the Centre d'Etude des Supports de Publicité (CESP). IREP and CESP perform in France many of the functions undertaken by ARF in the United States.**

## METHOD

Let:

a, b, c, . . . n be the vehicles of the combination, $A_a$, $A_b$, . . . $A_n$ be their respective audiences, and A the sum of these individual audiences. Thus,

(1) $A = A_a + A_b + \ldots + A_n$

Owing to duplication, the unduplicated audience C is smaller than the sum of the audiences of the vehicles, i.e.,

(2) $C = zA$ where
z is a coefficient between 0 and 1.

The proposed method consists in calculating z from the square table giving the duplication of media taken two by two. This table shows the audience duplicated by two vehicles at the intersection of their row and column: $D_{ab}$ is the number of people simultaneously covered by a and b.

| | a | b | c | . . . | n |
|---|---|---|---|---|---|
| a | — | $D_{ab}$ | $D_{ac}$ | . . . | $D_{an}$ |
| b | $D_{ba}$ | — | $D_{bc}$ | . . . | $D_{bn}$ |
| c | $D_{ca}$ | $D_{cb}$ | — | . . . | $D_{cn}$ |
| . | . | . | . | — | |
| . | . | . | . | | — |
| . | . | . | . | | |
| n | $D_{na}$ | $D_{nb}$ | $D_{nc}$ | . . . | — |

For obvious reasons this table is symmetrical about its main diagonal, D being the half-sum of all the terms in the table.

(3) $D = D_{ab} + D_{ac} + \ldots + D_{an} + D_{bc} + \ldots + D_{bn} \ldots$ etc.

Thus D represents the total of the two by two duplicated audiences. In the case of magazines, D represents the total of the double readerships. We have, on the other hand, the total duplicated audiences:

(1) $A = A_a + A_b + \ldots + A_n$

Suppose we let

(4) $x = \frac{D}{A}$

Now x is a factor which can be easily calculated starting from data already in our hands: the total audience of all vehicles and their two by two duplications.

We shall establish that there is a relation between x and the factor z of equation (2).

(5) $z = f(x)$

We shall prove that this relation may be satisfactorily represented by a unique curve the tracing of which will be given.

Knowing x, the value of z will be found on the curve, and this equation (2) will give a fair estimate of the unduplicated audience C, which is the object of our endeavor.

## THE RELATION BETWEEN x AND z

Obviously the greater the duplication between media, the less will be the unduplicated audience. The higher the value of D, the lower the value of C.

Inasmuch as $x = \frac{D}{A}$ and $z = \frac{C}{A}$, when x increases, z decreases.

When there is no duplication between media, the unduplicated audience equals the sum of the audiences. In this case, when $D = 0$, $x = 0$, and when $C = A$, $z = 1$.

Therefore when $x = 0$, $z = 1$.

Results released by the CESP enabled us to calculate the values of x and z for any combination of the 15 magazines for which the special study on unduplicated audiences was conducted.

To calculate x for a given combination, we divided the duplication between the magazines (as obtained from the two by two table) by the sum A of audiences of the magazines.

To calculate z, we divided the unduplicated audience C of the same combination (as actually observed in the special study) by the sum A of audiences of the magazines.

Theoretically we are in a position to work out 32,767 pairs of values for x and z. Actually we calculated only 98, selected at random from all combinations studied.

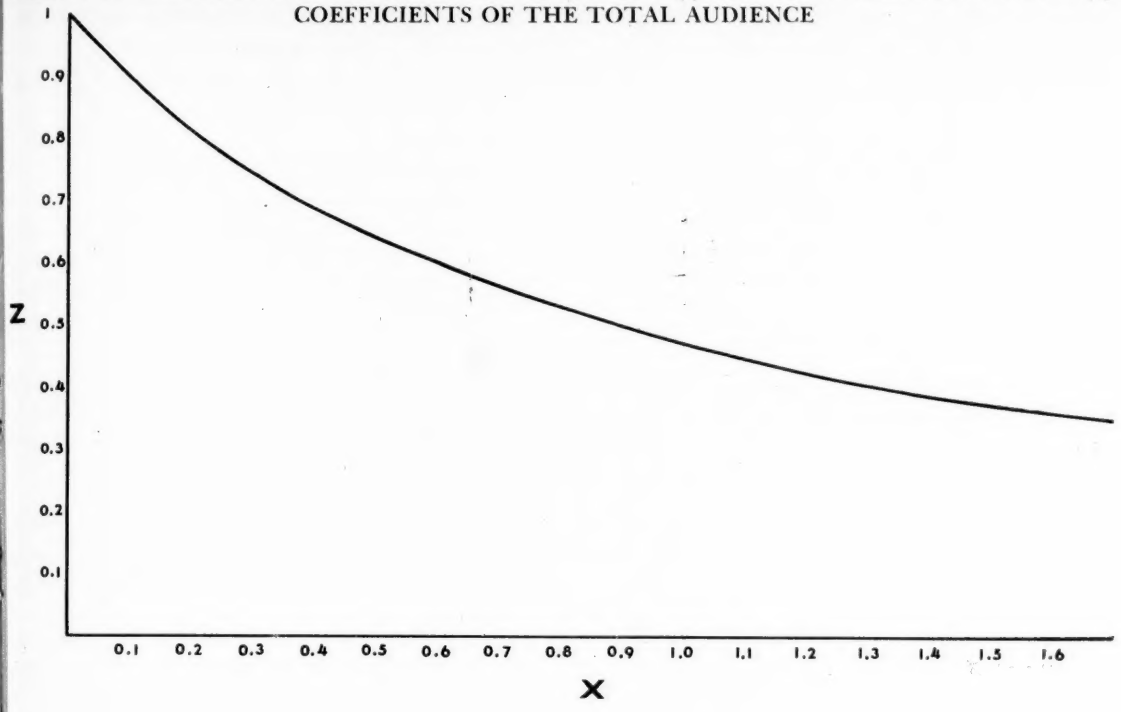
With x on the abscissa and z on the ordinate, we plotted 98 dots, the setting out of which clearly shows that the relation between x and z can be accurately depicted by a continuous curve (see Figure 1).

Such a curve was traced through the 98 dots. The accuracy of the graphic tool thus obtained was very satisfactory. The difference between the value of z read from the curve and the true value was in 90 per cent of the cases less than .01 and never more than .02.

We then tried to translate this curve into an equation. The tracing of the curve looks like an equilateral hyperbola asymptotic to the x axis. Obviously when vehicles are added to a combination the sum A may increase indefinitely while C must necessarily tend toward a limit, i.e., z must tend

FIGURE 1

OBSERVED RELATIONSHIP BETWEEN UNDUPLICATED (z) AND PAIR-WISE DUPLICATED (x) COEFFICIENTS OF THE TOTAL AUDIENCE




toward zero. The curve $z = f(x)$ must therefore have the x axis as its asymptote.

Accordingly we tried an equation of this form:

$$(6) \qquad z = \frac{1}{Kx + 1}$$

We found that when $K = 1.125$, equation (6) represented perfectly the section of the curve traced on the chart up to $x = 1.7$, the highest value of x observed in magazine combinations used to obtain the curve.

## PRACTICAL EXAMPLE

Let us, for example, take four magazines, the unduplicated audience of which has been calculated in the CESP special study. This will enable us to judge the accuracy of the proposed graphic method.

Suppose we have to estimate the unduplicated audience of the following four magazines, knowing the individual audience of each:

| | *Readers (in 000's)* |
|---|---|
| *Selection* | 4,741 |
| *Jours de France* | 1,573 |
| *La Vie Catholique* | 2,447 |
| *Nous Deux* | 4,143 |

We also know their two by two duplications:

| | *S* | *JF* | *VC* | *ND* |
|---|---|---|---|---|
| *Selection* | — | 638 | 663 | 697 |
| *Jours de France* | | — | 275 | 186 |
| *La Vie Catholique* | | | — | 283 |
| *Nous Deux* | | | | — |

We calculate

$$A = 4{,}741 + 1{,}573 + 2{,}447 + 4{,}143 = 12{,}904$$
$$D = 638 + 663 + 697 + 275 + 186 + 283 = 2{,}742$$

hence $x = \frac{D}{A} = 0.21$

On the curve we read the z value for $x = .21$, namely,

$z = .805$

hence $C = zA = .805\ (12{,}904) = 10{,}400$

The true value of C given by the CESP special study was 10,468.

## OTHER APPLICATIONS

The curve $z = f(x)$ on which the method is based was derived empirically from data on 15 French magazines. We have found that estimates obtained from the curve closely match the results actually

obtained in the CESP special study. But this is no proof that the curve is valid for other media than those which served to build it. We therefore decided to test the proposed method beyond those 15 magazines.

The CESP survey tabulated the unduplicated audience of a few groups of media other than the 15 chosen magazines. When the two by two duplications are given as well, it is possible to apply our method and compare the estimate obtained with the actual result. We conducted three tests with:

a. five women's magazines (of which only one was part of the 15);

b. five general magazines (of which none was part of the 15);

c. eleven Paris dailies.

In the three cases, the proposed method has led us to a fairly close estimate of the unduplicated audience:

| | *Estimated Audience* | *Actual Audience* | *Difference* |
|---|---|---|---|
| a. | 5,700,000 | 5,837,000 | 2.3% |
| b. | 6,200,000 | 6,210,000 | 0.2% |
| c. | 6,000,000 | 6,160,000 | 2.6% |

All media previously mentioned are French publications the audiences of which were measured by this same CESP survey. It seemed advisable to test the proposed method with U. S. magazines. This we were able to do thanks to the 1954 *Life* report, *A Study of Duplication.* In this study, audience duplications of five U. S. magazines *(Life, Saturday Evening Post, Look, Ladies' Home Journal, This Week)* were analyzed.

Specifically, unduplicated audiences were given for all combinations of three, four and five magazines. In each case these audiences were analyzed according to sex and income. We thus have a total of 96 results regarding unduplicated audiences—16 combinations for each of six groups (one over-all, two sexes and three income levels). The same audiences were evaluated by applying our method with the same constant, K = 1.125. Results:

In 82 cases, the difference between the actual audience and the estimated audience was less than one per cent.

In ten cases, this difference lay between one and two per cent.

In four cases, this difference exceeded two per cent (maximum 2.7 per cent).

It appears that the proposed method has a wide field of application.

## CONCLUSIONS

If it is true that the only information required to estimate an unduplicated audience is the total audience and two by two vehicle duplications, then great economy of effort is possible. Instead of undertaking the fantastic amount of tabulation required to calculate the unduplicated audiences of all possible combinations of a restricted number of vehicles, it is preferable to set up, for all vehicles, a series of square tables of their two by two duplications, each table containing these duplications within one stratum. It should be noted that to evaluate the unduplicated audience of a combination of vehicles within a stratum (e.g., an age group), one must know the duplications *within this stratum.* Knowledge of duplications for the whole sample is not sufficient.

We feel that an analysis of the two by two duplications by strata for any individual audience analysis is much more efficient than a repetition of such voluminous work as the CESP special study. We must not forget, however, that if this tremendous enterprise had not been undertaken, we would not have found a method to save us from doing it in the future.

*General observations drawn from particulars are the jewels of knowledge, comprehending great store in a little room.*

—JOHN LOCKE

# A Switch-Over Experimental Design to Measure Advertising Effect

R. J. JESSEN
University of California at Los Angeles
and C-E-I-R, Inc.

**The influence of advertising may endure long after its placement. Dr. Jessen describes an experiment which takes account of such delayed effects while measuring sales response to current advertising outlay.**

AN IMPORTANT GOAL in determining the size of an advertising budget is to maximize profits. To achieve this goal, even approximately, is a very complex problem. Many procedures or rules have been adopted to provide what appears to be a reasonable answer from one point of view or another, but which, most people in the business would agree, are very crude at best. Hence, such approaches as "percentage-of-sales," "all-you-can-afford," "objective-and-task" and "competitive-parity" are used because "scientific" approaches either don't exist, or have not been able to prove their superiority.

The difficulties of devising a scientific approach are fairly well known. The multitude of factors involved and the scarcity of relevant and accurate data are only two of many which could be listed. But work is being done to develop good methods and, with the increasing interest that researchers are paying to the problem, there is no doubt that real progress is being and will be made.

The purpose of this paper is to present a scheme for determining the sales response to varying advertising outlays for a very simple case. It is hoped that the method can be extended to more complex and therefore more realistic cases once the underlying principles are thoroughly understood.

Although we are concerned here with the response of sales to different amounts of advertising outlay, it should be relatively straightforward, at least in principle, to proceed to the determination of the optimum outlay.

Suppose we represent the general relationships existing between the advertising outlay of a firm, its consequent sales and profits as those shown in

**RAYMOND J. JESSEN is lecturer in business statistics at UCLA and project director at C-E-I-R, Inc. in Los Angeles. He received a B.S. in economics from the University of California in 1937 and a Ph.D. in statistics from Iowa State University in 1943. Until 1947, he was an agricultural statistician at the U. S. Bureau of Agricultural Economics. For the next ten years he taught statistics at Iowa State University while consulting on survey techniques with the Rockefeller Foundation, United Nations, Tennessee Valley Authority and U. S. Departments of Agriculture and the Interior. His publications in scientific journals have emphasized gauging the cost and effectiveness of sampling techniques. He is a fellow of the American Statistical Association and a member of the American Marketing Association, Biometric Society and Inter-American Statistical Institute. He also serves on the A.S.A.'s three-man Technical Committee on Broadcast Ratings.**

FIGURE 1
GENERAL RELATIONSHIP OF REVENUE AND PROFITS TO ADVERTISING OUTLAYS

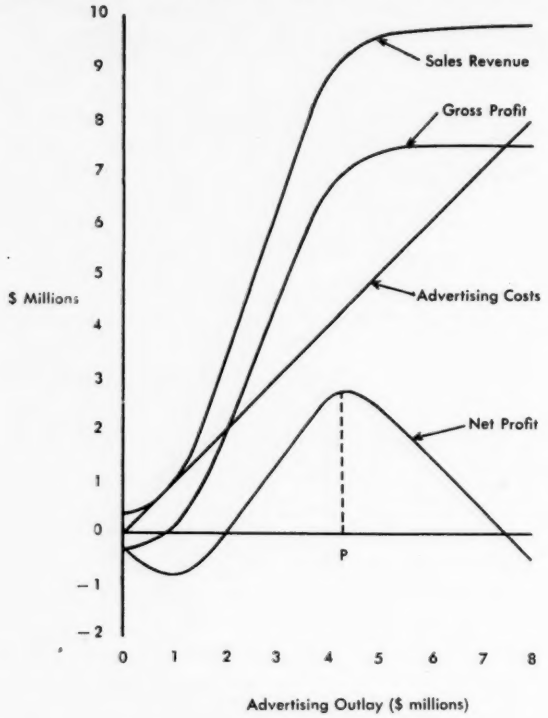



Figure 1. To keep our model simple, let us assume that the "best" use is made of the advertising budget at each outlay level and that disturbances such as seasonal variations, general state of the market and the economy, competitors' activities, etc., are held constant.

Sales are expected to increase, but not in proportion to advertising expenditures. Increasingly heavy doses of advertising are expected sooner or later to meet with a diminishing increase in sales. If gross profits are defined as simply Sales Revenue less all costs except those for advertising, we obtain the Gross Profit curve (which may be negative for low advertising outlays). Next, a curve is shown representing increasing Advertising Costs, and lastly, the Net Profit curve which is simply the Gross Profit less Advertising Costs. Net Profit is a maximum at Point "P" where $2.75 million net profit is made with an advertising budget of $4.25 million. Any other size of budget will yield less profit.

This paper is concerned with a procedure for determining the top curve—the response of sales to total advertising outlays. The use of the budget among the different media and the particular kind of copy presented need not be fixed. The conditions which must be satisfied will be mentioned later.

The remainder of the paper will present:

1. a brief statement on basic statistical requirements for planning and conducting experiments;
2. special problems in dealing with advertising budgets;
3. a layout for a switch-over type design;
4. some methods of estimating "carry-over" effects;
5. the results of a hypothetical experiment which estimate these effects;
6. discussion of the applicability of this experimental design.

## METHOD

### *Experiments*

A controlled experiment is an investigation in which the experimenter chooses *which* treatment to apply *when* and *where*, and then observes the response. The design and analysis of complex experiments is a branch of applied mathematics largely originating in the biological sciences and in recent years is gradually appearing in a number of other sciences. Its penetration into economics and studies of human behavior has been particularly slow—partly because of the complexity of many of the problems of these fields, and partly because of unfamiliarity with modern developments in modern experimental techniques.

It was a long-held scientific doctrine that good experimental procedure was to study one variable at a time holding all others constant. Thus, one might study the relationship of the occurrence of stomach ulcers in monkeys to kinds, intensity and frequency of stress by varying the intensity while holding kind and frequency constant. Having determined this relationship, one would vary frequency while holding kind and intensity constant and so on. Such a procedure is not only quite wasteful of time and resources, but provides insufficient information. If there is an interaction between intensity and frequency, it would never be discovered by this method. The correct method is to determine the effects of intensity, frequency and kind, and their interactions *simultaneously* in a single well-planned experiment.

To design experiments appropriate for a given investigation requires some familiarity with experimental design as an art and science, and some of the relevant lore of the field in which it is being applied. If the problem is such that standard techniques, even with modification, will not apply, it may be advisable to devise new procedures.

### *Advertising Budgets*

There are a number of special problems that immediately become apparent when one considers good experimental procedure for measuring the effects of advertising budgets. Often these problems have made some would-be experimenters shy away with the remark that experimentation may be all right for agriculture in testing for yield effects from different fertilizer treatments or in engineering research, but it just can't be used in the social sciences.

Listed below are some of these problems, with comments:

1. What is a "dose" of advertising expenditure? The amount of dose can be varied by changes in frequency, coverage, space units (e.g., color versus black and white), copy, etc.
2. Carry-over effects. After an advertising treatment is removed, there is still a residual effect left which will continue to influence sales for awhile.
3. Seasonal effects. Sales may have a pronounced seasonal pattern and it may be necessary to remove this influence. Also, a given dose of advertising given in the spring may bring a response different from the same dose given in the fall.
4. Geographical effects. Responses to advertising differ from one locality to another due to income structure, influence of competitors' position, etc.
5. Competitors' programs. If a competitor changes his advertising during the course of an experiment it may confound results.
6. Shifts in consumers' behavior. If the experiment takes a long time to complete, basic consumer behavior in regard to advertising may change.
7. Other factors affecting response. Changes in weather, the world political and economic situation, etc., during an experiment may cause trouble.
8. Measuring sales. Records on sales may not be available for experimental markets or for the time periods desired. Perhaps there have been price changes in the product, or a new premium scheme has been launched.
9. Interference of the experiment with going operations.

A properly designed experiment would have to take problems of the sort listed above into account or else it is doomed to failure.

### *A Simple Switch-Over Design*

Consider the following illustrative problem. A company has sales outlets in 85 market areas into which it divided the U.S. (to make subsequent uses here simple), and its advertising is carried out locally in each of its market areas. Suppose its advertising budget in each market area is now $1,000 per month. What would be the sales response to budgets of $2,000 per month? To $3,000 per month? These outlays would correspond roughly to annual expenditures of one, two and three millions (see Figure 1).

Denote test areas by 1, 2, 3, etc., time periods by I, II, III and budget levels by A, B, C, where the lower budget rate (A) is $1,000, the medium (B) is $2,000 and the high (C) is $3,000.

We wish to put the treatments A, B and C on *each* area during the experimental period in order to eliminate area differences in our comparisons. It is reasonable to believe that the *order* of treatments will affect results; that is, the sales response in each period to an advertising sequence of ABC will likely be different from that of CBA. (This would mean a $1,000-$2,000-$3,000 sequence as contrasted with a $3,000-$2,000-$1,000 sequence.) Since there is a total of six possible sequences, a minimum of six areas will be required for the experiment. These can be arranged into two groups of three areas each, the areas having similar seasonal patterns being put together into some group. With this design, each group can be regarded as a separate experiment known as a 3 x 3 latin square. There will be two basic types to cover all possible sequences (see Figure 2).

FIGURE 2

3 x 3 LATIN SQUARE & ITS COMPLEMENT

| Period | Area 1 | Area 2 | Area 3 |
|---|---|---|---|
| I | A | B | C |
| II | B | C | A |
| III | C | A | B |

| Period | Area 4 | Area 5 | Area 6 |
|---|---|---|---|
| I | A | B | C |
| II | C | A | B |
| III | B | C | A |

For our purposes it may be better to combine the two squares into a single experiment with rows (periods) regarded as constant throughout the U.S. In this case the breakdown of degrees of freedom for the purpose of determining 'experimental error" would be as follows:

| *Source of Variation* | *Degrees of Freedom* |
|---|---|
| Treatments | 2 |
| Periods | 2 |
| Sequences | 5 |
| Residuals | 2 |
| Error | 6 |
| Total | 17 |

In a later section, suggestions are given for the appropriate method of computing the mean squares of this table for this design.

### *Carry-Over Effects*

The switch-over type design can be useful in eliminating errors due to area differences, both in the average main effects of the budget treatments and in the average seasonal effects, by balancing one effect against another. If carry-over exists, the unadjusted observed effects will be biased. Adjustments will be required if estimates free of bias are to be made of what the sales effect for various budgets would be if they were carried out for extended periods of time.

The nature of carry-over effects in the proposed experiment may be illustrated by extending our example. Suppose our company, whose advertising sales-profit functions are shown in Figure 1, carried out an advertising budget experiment on a set of six test market areas chosen by some randomized scheme. A switch-over type design was used where three levels of budgets were run in the six possible sequences shown in Figure 2. A response to a given budget level is read off the response function in Figure 1. For example, if the budget in a given market area is $1,000 per month, it would correspond to a national annual level of $1,000,000 and would produce an expected response of $1,000,000 in sales at the national annual level, or $1,000 per month in the individual market area. Likewise, budgets of $2,000 and $3,000 per month would produce expected sales responses of $3,400 and $6,300 per month, respectively.

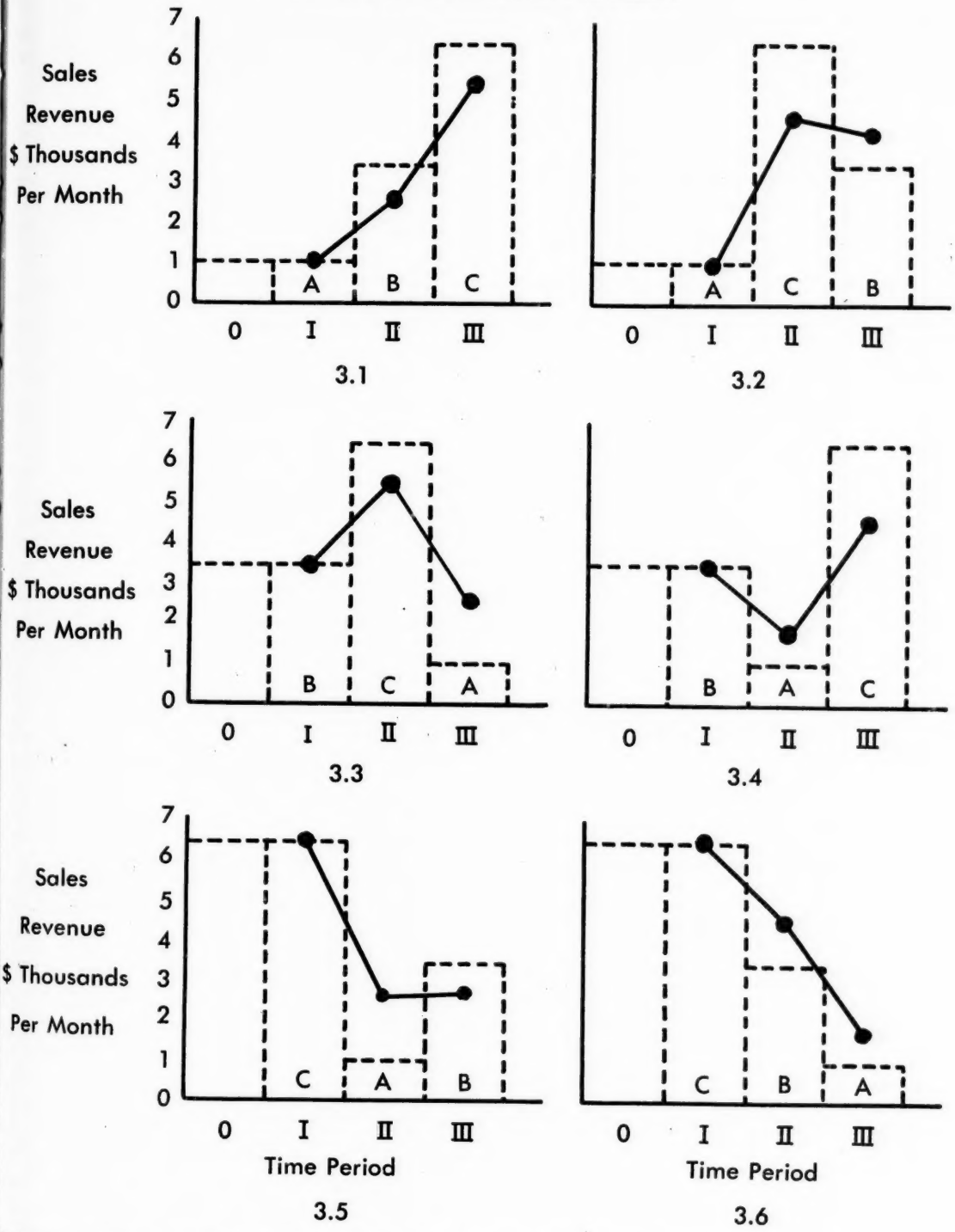
In Figure 3, the response to each budget is shown for the two cases: (1) with and (2) without response lag or carry-over. The broken lines indicate the response for each period if there were no lag; the small circles (connected by solid lines) the average response for the period considering that some sales may lag somewhat behind the input of advertising effort. It is assumed that experimental periods for a trial run at a given level are so short that some sales triggered by advertising in one period occur during the subsequent period when a new level is being put to test. It is being presumed that the advantages of a short period are great enough to make it worthwhile to complicate the statistical task that this design entails.

The general effect of lags is to obscure the underlying relationships. For example, in box 3.1 of Figure 3, the test area is presumed to have been run at the $1,000 budget level (A) for the period before the experiment is regarded underway. The sales response in period I in both the case of no-lag and lag-response, is the same—that is, $1,000 per month. In period II, instant responses would provide sales of $3,400 per month, but with lag, the observed response is $2,680. Likewise, in period III the values are $6,300 and $5,430, respectively. In box 3.2 note that lag reduces the sales in period II, but increases it in III when the budget was reduced from the high to a medium level.

To look at the nature of response of sales to an advertising stimulus, let us consider a very simple model. Suppose we consider what happens if we put a small "dose" of advertising into a market. Following the advertising, we might expect that sales would rise to some peak and then trail off. We might represent this relationship by the simple graph at the top of Figure 4, where a dose of advertising was applied in the middle of period I and sales increased from zero to a maximum by the end of the period, two-fourths later, then trailed off to zero again in another two-fourths of a period. The total sales resulting from this dose of advertising is represented by the area under the curve (or the sum of the columns). This is the total response. However, only half the sales occurred in the period when the advertising was carried out. A "residual" of sales carried over to the subsequent period.

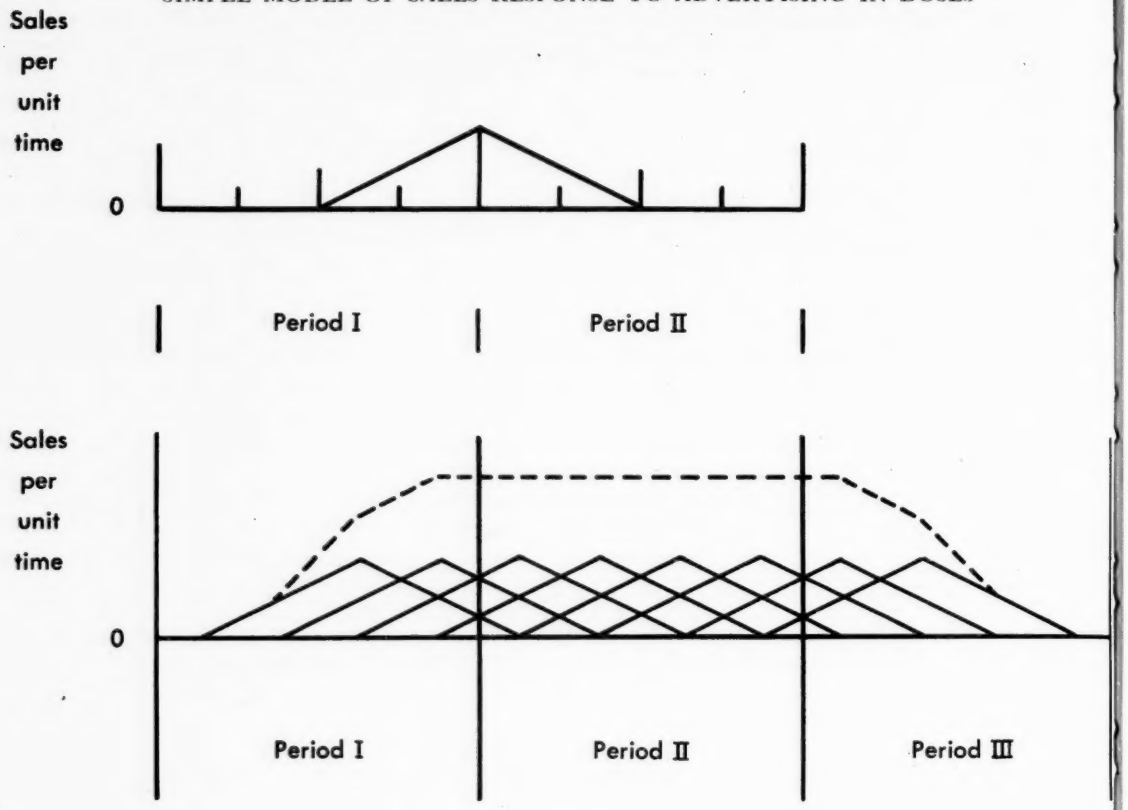
If we consider a series of doses during a given period and let each respond in this way independently of each other, we have a picture like that at the bottom of Figure 4. Advertising outlays made during period I will build up to a maximum (in this case by the end of the period), and if the outlays are continued through period II, no further gain is made. Hence, in period II, a kind of equilibrium is reached where sales are at a maximum for this level of advertising outlay. If the advertising is withdrawn in period III, the sales will trail off to zero again. Later, we shall denote the sales in the first period as D, the direct response; those in the second period as T, the total response; and those in the third period as R, the residual response.

FIGURE 3
ILLUSTRATIVE RESPONSE PATTERNS




Broken line histograms indicate response assuming no carry-over. Small circles indicate observed response assuming carry-over.

FIGURE 4
SIMPLE MODEL OF SALES RESPONSE TO ADVERTISING IN DOSES




## RESULTS

To provide illustrative material for Figure 3, a model was developed to represent the lag or carry-over of some advertising. Suppose we represent the amount of sales response observed in a given market, during a given period and at a given budget level, as a sum of a number of component effects:

$$y_{ijk} = D_k + R_k' + a_i + p_j + e_{ijk}$$

where,

$y_{ijk}$ = the observed response in the $i^{th}$ market area during the $j^{th}$ period when treatment k is being applied following treatment k′,
$D_k$ = the actual response to treatment k during its period of application, or "direct" response,
$R_k'$ = the actual response to treatment k′ applied to the preceding period occurring during the present period, or "residual" response,
$a_i$ = the differential effect of the $i^{th}$ market area,
$p_j$ = the differential effect of the $j^{th}$ period, and
$e_{ijk}$ = all other effects, to be regarded as residual "error."

Since we are interested in the total sales effect of a given advertising application, we need to determine both D and R for a given treatment; that is, how much we get currently and how much subsequently. If we denote total response of treatment k by $T_k$, then

$$T_k = D_k + R_k$$

The switch-over type of design, with modifications, will permit unbiased estimates to be made of the total responses, $T_k$. As an illustration of the kind of data required and the computations involved, let us continue with the example used so far. For the true values of $T_k$ we have for treatment levels A, B and C:

$T_A = 1.0$ (thousands of dollars per month)
$T_B = 3.4$ (thousands of dollars per month)
$T_C = 6.3$ (thousands of dollars per month)

Now, let us assume that direct and residual portions of the total responses to advertising are:

$$D_k = w\,T_k$$
$$R_k = (1 - w)\,T_k$$

where w is simply the fraction of the total response occurring during the period of the treatment, and (1 — w) is the remaining fraction occurring in the subsequent period. If we assume for our example that w = .70 then,

Treatment

| k | $T_k$ | $D_k$ | $R_k$ |
|---|---|---|---|
| A | 1.0 | .70 | .30 |
| B | 3.4 | 2.38 | 1.02 |
| C | 6.3 | 4.41 | 1.89 |

The observed responses for each of the six sequences of the three treatments is simply

$$y_k = D_k + R_k'$$

where $R_k'$ is the residual from the previous treatment. The $y_k$'s for our example have been determined and plotted in Figure 3.

Let us assume that the period effects are:

$p_1 = +0.1$ (thousands of dollars per month)
$p_2 = 0.0$ (thousands of dollars per month)
$p_3 = -0.1$ (thousands of dollars per month)

This means that during period I the sales response for $1,000 per month of advertising outlay will be $100 per month more than the average period; whereas, in period III, it will be $100 per month less.

For market area effects we have (in thousands of dollars per month)

$a_1 = -0.3$ $\quad a_4 = +0.1$
$a_2 = -0.2$ $\quad a_5 = +0.2$
$a_3 = -0.1$ $\quad a_6 = +0.3$

That is, market area 1 during any period always responds with $300 less sales per month for $1,000 advertising outlay than the average market area; whereas market area 2 is $200 less, etc.

Assuming residual errors are zero and aggregating the effects from the several sources, we have in Table 1 the values which would be observed in our switch-over experiment with the six test market areas. (Figure 3 is based on these data, except that area and period effects were omitted.)

TABLE 1
"OBSERVED" SALES IN THE SWITCH-OVER EXPERIMENT

| | *Market Area* | | | |
|---|---|---|---|---|
| *Period* | *1* | *2* | *3* | *Total* |
| 1 | A = 0.80 | B = 3.30 | C = 6.30 | 10.40 |
| 2 | B = 2.38 | C = 5.23 | A = 2.49 | 10.10 |
| 3 | C = 5.03 | A = 2.29 | B = 2.48 | 9.80 |
| | 8.21 | 10.82 | 11.27 | 30.30 |
| | *4* | *5* | *6* | *Total* |
| 1 | A = 1.20 | B = 3.70 | C = 6.70 | 11.60 |
| 2 | C = 4.81 | A = 1.92 | B = 4.57 | 11.30 |
| 3 | B = 4.27 | C = 4.81 | A = 1.92 | 11.00 |
| | 10.28 | 10.43 | 13.19 | 33.90 |

It may be of interest to see if the basic response data fed into the system can, in fact, be recovered. To do this, calculations of the following quantities are required:

- $H$ = total responses for all period-areas on which treatment k was put. This is the unadjusted total for each treatment.
- $F$ = total responses for all period-areas immediately following the one on which treatment k was put (including the pre-experiment treatment).
- $S_L$ = total of sequences where treatments occurred last.
- $S_F$ = total of sequences where treatments occurred first.
- $G$ = grand total of the experiment.
- $y$ = grand mean of the experiment.

The computed values of these quantities are shown in Table 2.

To obtain an estimate of $T_k$, the total sales attributable to the $k^{th}$ advertising budget, both direct and residual, will be given by the estimator,[1]

$$\hat{T}_k = \frac{H}{6} + \frac{1}{12}(3F - G + S_L - S_F)$$

[1] The author is indebted to Dr. M. R. Mickey for this formula and for the appropriate method for computing "experimental error" for this design.

For treatment A, the budget rate of $1,000 per month, we obtain as an estimate of total sales,

$$T_A = \frac{10.62}{6} + \frac{1}{12}[(3)(16.48) - 64.20 + 24.01 - 18.49]$$

$$= \frac{10.62}{6} - \frac{9.24}{12}$$

$$= 1.00$$

Hence, we have "recovered" the correct value of treatment A from our dummy experiment. In a similar manner the correct responses for treatments B and C can be estimated.

If we define the differential direct and residual effects in the system as

$$d_k = D_k - \bar{D} \text{ and } r_k = R_k - \bar{R}$$

respectively, where $\bar{D}$ and $\bar{R}$ are the means of the component effects respectively, then the true values for these quantities in this example are:

| Treatment | $d_k$ | $r_k$ |
|---|---|---|
| A | −1.80 | −0.77 |
| B | −0.12 | −0.05 |
| C | +1.92 | +0.82 |

and $\bar{D} = 2.50$, $\bar{R} = 1.07$ and $\bar{T} = 3.57$. To estimate $r_k$ we have

$$\hat{r}_k = \frac{1}{12}[3F - G + S_L - S_F]$$

and

$$\hat{r}_A = \frac{1}{12}[(3)(16.48) - 64.20 + 24.01 - 18.49] = -0.77$$

To estimate $d_k$ we have

$$T_k = \bar{T} + d_k + r_k$$

hence $\hat{d}_k = \hat{T}_k - y - \hat{r}_k$

and $\hat{d}_A = 1.00 - 3.57 + 0.77 = -1.80$

where $\bar{T}$ is estimated by y, the grand mean of the experiment. In general, we cannot estimate $\bar{D}$ and $\bar{R}$ individually with this design, but it is not of importance here anyway.

It should be noted that certain properties are peculiar to this design. The treatments are applied during the pre-experimental period so that observations made during the first period include residuals of known type. This also permits a simple method of estimating what is important here, the total responses $T_k$ as well as the components, $d_k$ and $r_k$. An estimate of w is also available from the experiment itself.

The variance of $T_k$ is given by

$$\hat{\text{Var}}(\hat{T}_k) = \frac{\sigma^2}{3}$$

where $\sigma^2$ is the appropriate experimental error for this design and

$$\hat{\text{Covar}}(\hat{T}_A, \hat{T}_B) = -\frac{\sigma^2}{12}$$

TABLE 2
COMPUTATION SUMMARY OF TOTAL, DIRECT AND RESIDUAL EFFECTS

| Treatment | H | F | $S_L$ | $S_F$ | $\hat{T}_k$ | $\hat{d}_k$ | $\hat{r}_k$ |
|---|---|---|---|---|---|---|---|
| A | 10.62 | 16.48 | 24.01 | 18.49 | 1.00 | −1.80 | −0.77 |
| B | 20.70 | 21.10 | 21.55 | 21.25 | 3.40 | −0.12 | −0.05 |
| C | 32.88 | 26.62 | 18.64 | 24.46 | 6.30 | +1.92 | +0.82 |
| Totals | 64.20 | 64.20 | 64.20 | 64.20 | 10.70 | 0.00 | 0.00 |

$G = 64.20$
$y = 64.20/18 = 3.567$

The appropriate method of computing the sums of squares for the analysis of variance table given in the previous section is conventional for all sources except residuals. Here the sum of squares is

$$4\,(\hat{\tau}_A^2 + \hat{\tau}_B^2 + \hat{\tau}_C^2)$$

The rather serious bias which can arise if the data are not treated properly, and which obscures the true relationship of sales to outlay, may be examined in this case. The unadjusted treatment totals are given by the T's in the table below. From these we can compute the unadjusted treatment means. Thus, for the biased estimates of $T_k$ we have,

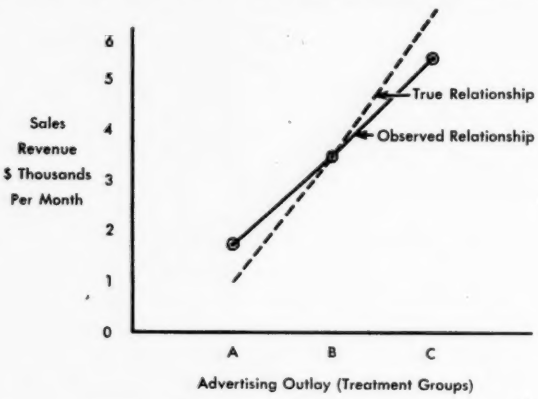
$T_A' = 10.62/6 = 1.77$

$T_B' = 20.70/6 = 3.45$

$T_C' = 32.88/6 = 5.48$

FIGURE 5

DISTORTION FROM CARRY-OVER EFFECTS




When plotted against the true values (see Figure 5) the nature of the bias becomes apparent. In this case the response to low levels of outlay are overestimated and the high levels are underestimated. If guided by such results one may miss the true optimum level by a wide margin.

## DISCUSSION

The several steps still remaining will only be briefly mentioned here and perhaps dealt with in some detail in a later paper. The next step in this plan is to compute the Net Profit curve from the estimates of the portion of the Gross Sales curve covered by the experimentation (Figure 4), and from estimates of production costs. An examination of the information should be helpful to determine where the advertising budget should be set for all non-experimental market areas, and what levels should be chosen for further experimentation.

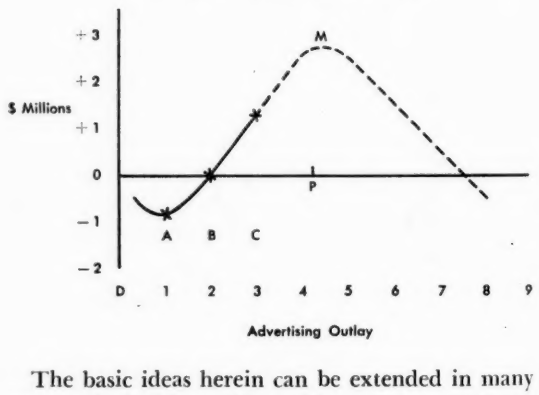
Note in Figure 6 that our illustrative experiment explored only a small portion of the Net Profit curve. Future experimentation would test higher levels of advertising outlay in order to seek the maximum at M. (Bear in mind that the experimenter would not know the portion of the curve shown by a broken line.) He may be quite cautious, thinking that level C is near maximum already. The strategy to use at this point could be an interesting study for the future.

It is foolish to think that these relationships are fixed and once determined, will form the basis for all future action. Rather, it is more reasonable to regard the relationships as undergoing constant changes and that a business organization is faced with the problem of constantly seeking out the relationships *currently* in effect so that it can behave most rationally in regard to them. This requires that it keep up constant experimentation to determine where these curves are—perhaps, some day, to forecast them.

Some businesses will immediately see that a plan such as this one requires that sales be known by both market area and period. This will be a serious limitation unless measures are taken to obtain such information either from commercial services or by organizing their own. The advisability of either depends on expected gains against costs.

FIGURE 6

THE NET PROFIT CURVE




The basic ideas herein can be extended in many ways to deal with specific problems. For example, in a manner similar to that presented, an experimental plan could be made which would evaluate the effectiveness of different media rather than budget levels, or a combination of both. A more elaborate design could evaluate both media and budget levels. It appears that the response curves for different levels of inputs in the several media may differ somewhat, but what is probably more important are the interactions of media. Certain media mixes may have a total effect different from what the sum of the component parts might indicate. These differences can also be isolated and measured by appropriately designed experiments.

# The Influence of Newspaper Tune-In Advertising on the Size of a TV Show's Audience

JAMES C. BECKNELL, JR.
E. I. du Pont de Nemours & Company

**Twenty-one cities were randomly assigned to groups receiving normal, double and no tune-in advertising for a special TV program. Coincidental viewing measures indicated that the ads had no effect on audience size, but may have served to decrease audience variability.**

THE ADVERTISER who decides to hawk his wares on national television is committing himself to the expenditure of considerable sums of money. Anything he can do to increase the size or quality of the audience should be evaluated by him, and accepted or rejected on a sound cost basis. The advertiser spending half a million dollars to buy the attention of 20 million people would be foolish if he did not spend an additional ten thousand for one million more. However, the advertiser spending only 100 thousand for the same 20 million people might be foolish to spend an additional ten thousand for a five per cent audience increase.

This, of course, assumes that the probability of purchasing (or behaving in some other desired manner) is the same for the last million as it was for the original 20 million—a completely ridiculous assumption. Obviously some people are far better targets for an advertising message than others. One may ask whether advertising is more efficient when it selects the "right" people rather than when it gathers a large number of viewers indiscriminately. This study does not attempt to answer this question.

Rather, it asks: after a given expenditure for time and talent, what can an advertiser do to increase viewing? Several avenues are open. He can improve the quality of his program (assuming that the expenditure of additional monies will do this). He can advertise his program in spot television messages, in newspapers, in magazines or anywhere he chooses. The purpose of this study is to evaluate only one of these possible methods for increasing audience size, advertising the show's presence on the television page of newspapers the day of the show.

The logic behind promoting a special television program seems infallible. People can't get into a regular habit of watching such a program so they must be made aware of it. Awareness comes either

**JAMES C. BECKNELL, JR. is a research psychologist in the Advertising Research Section of E. I. du Pont de Nemours & Company. He has worked for R. H. Macy as an executive trainee and assistant buyer, and for the A. J. Wood Research Corporation as a project director. He received a B.S. in economics from the University of Pennsylvania in 1952, an M.A. in general psychology from the University of Arkansas in 1957, and is presently completing work for the Ph.D. in psychology at the University of Pennsylvania. He is a member of the American Psychological Association and the American Statistical Association.**

through their own volition (actively searching television listings) or by having it thrust upon them. The latter can happen either by their accidentally tuning in the program, or by noticing some form of announcement that the program will be shown at a given time. The same reasoning may hold partially for regularly scheduled shows, but it is not so compelling.

The *post hoc* experiment or analysis of historical data is one method often used for evaluation of a promotional technique. At best, *post hoc* experimentation can yield only hypotheses to be tested later in a completely designed study. This is especially true in the case of unusual events such as TV spectaculars which may differ widely in appeal. Whether a spectacular was or was not advertised depended on a variety of reasons. If there were any reason other than random assignment of shows to the categories, "promoted" and "not promoted," *post hoc* analysis is inappropriate.

For this reason, the decision was made to conduct a field experiment to determine the effect of newspaper tune-in advertising on the Du Pont "Show of the Month" run in April 1960.

## METHOD

Twenty-four cities where tune-in advertising is normally run for the "Show of the Month" (on the television page of the newspaper the day of the show) were selected and stratified according to newspaper circulation. Three of these cities were eliminated: New York and Chicago, where the pattern of advertising differed from that of the other cities, and the Philadelphia-Wilmington area because the Du Pont home office is located there. The rationale behind stratifying on newspaper circulation rather than size of city was the desire and appropriateness of equalizing cost. Stratification could have been accomplished by the cities' former viewing pattern of the show but this was rejected because of within-city variance.

After the stratification, cities were grouped by threes and randomly assigned to one of three experimental conditions. Group A, which received the normal size tune-in advertisements (300 lines), contained the following cities:

Detroit
Minneapolis-St. Paul
Pittsburgh
Kansas City, Mo.
Milwaukee
Cincinnati
Columbus

Group B, which received no advertising, contained the following cities:

Boston
Atlanta
Washington, D. C.
Cleveland
Denver
Buffalo
Seattle-Tacoma
Houston

Group C, which received a double size advertisement (600 lines) contained the following cities:

San Francisco-Oakland
Baltimore
St. Louis
Miami
Indianapolis
Dallas-Ft. Worth

Because of the random process used for assignment, Minneapolis-St. Paul and San Francisco-Oakland were separated into different groups. Because of the newspaper overlap in these areas, San Francisco and St. Paul were swapped to obtain geographical harmony, and the ensuing pairs were randomly reassigned. These pairs are counted only as one observation.

On April 21, 1960, the day of the Du Pont "Show of the Month"—"Years Ago"—an advertisement (see Figure 1) was run in the newspaper which the company's advertising department and agency normally considered to be the best buy in each city. That evening the American Research Bureau of Beltsville, Maryland made approximately 375 telephone calls during the "Show of the Month" in each of the 21 cities. An average of 66 per cent of the calls were completed in each city. The standard deviation around this 21-city mean was 4.7 per cent. The data of this study are based on these interviews. Previous ARB data on the November, December, January and March "Show of the Month" were utilized in making statistical adjustments and in determining the over-all variance estimate in the analysis.

Two hypotheses were specifically tested. They were stated in the null form:

1. There will be no differences in the mean viewing levels among cities which had no tune-in advertisements, 300-line tune-in advertisements, or 600-line tune-in advertisements.
2. There will be no difference in the variance estimates for the groups of cities which had no tune-in advertisements, 300-line tune-in advertisements, and 600-line tune-in advertisements.

## RESULTS

Table 1 shows the mean proportions of television sets (of all television sets) reported to be tuned in to the April "Show of the Month" by ARB's coincidental interviews. Also shown is the mean value for the previous four shows' average score. There were no significant differences among the experimental treatments either during the test show or during the preceding periods. But the difference between the April show and the average of the previous shows was significant. It is interesting to note that the slight superiority of Group C (double advertising) in April is even more pronounced in the averages for the previous four shows when all groups received the same advertising. As shown in Table 1, ratios of the April ratings to the average ratings from the previous four shows bring out even more strikingly that there was no effect of newspaper tune-in advertising on total audience size. This was true not only of the proportion of all sets as shown, but also of the proportion of *sets in use*.

TABLE 1

MEAN PER CENT CLAIMED VIEWING

| *City Group* | *April* | *Four Shows' Average* | *April ÷ Average* |
|---|---|---|---|
| B | 13.7 | 16.5 | .817 |
| A | 13.8 | 18.3 | .757 |
| C | 15.6 | 19.1 | .822 |

In April cities in Group B received no tune-in advertising; Group A received regular advertising; Group C received double the regular advertising. The average ratings of the three groups in April did not differ significantly from one another. All three groups, however, received significantly lower average ratings in April than in the previous four months, when regular tune-in advertising had been applied to all of them.

To test the second hypothesis, the decision was made to compare the variance from each of the experimental groups against the best over-all estimate we had of variance for the population of cities. This was done by testing the variance estimate of each subgroup against the over-all variance estimate for the population of shows included in the average scores for the preceding four shows, plus the "normal" advertising condition for the April show. This test found that the variance between cities within the experimental group that received no tune-in advertising was significantly greater than the variance normally expected in this population. This same tendency was also seen in the other two experimental groups. However, because of a lack of degrees of freedom, it cannot be said to be significant.

FIGURE 1

THE TUNE-IN AD TESTED

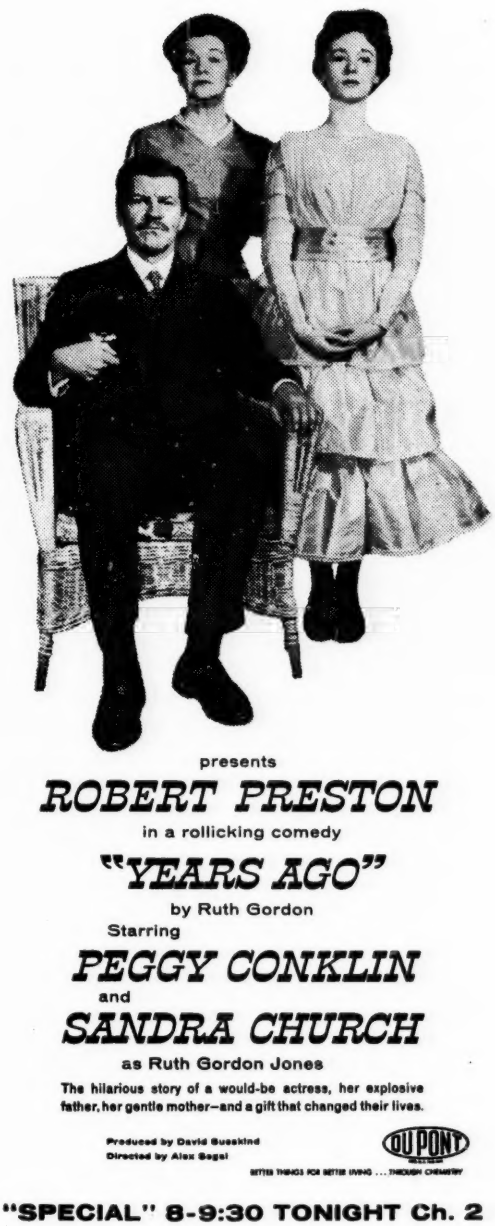


TABLE 2

VARIANCE OF RATINGS AMONG CITIES

| *Advertising Level* | *No. of Cities* | *Variance* |
|---|---|---|
| None | 8 | 39.5 |
| Regular | 7 | 22.5 |
| Double | 6 | 13.6 |

The variance of ratings among cities with no tune-in advertising was significantly ($P > .99$) greater than the variance (11.5) among cities with regular advertising. To calculate this variance of cities with regular advertising we have considered all 21 cities for each of the previous four shows as providing independent ratings. To them we added the ratings of the seven cities which received regular advertising in April.

The "Show of the Month" began at 8:30 p.m. in some cities and 9:30 p.m. in others. The average rating for the 9:30 group (16.5) was significantly greater than that for the 8:30 group (9.4). The various advertising conditions, however, were fairly well distributed over the two time zones.

## DISCUSSION

On this evidence hypothesis one cannot be rejected. That is, there were no data which indicated any relationship between tune-in advertising and the level of television viewing.

Hypothesis two, however, must be rejected. It was a surprising finding that the variances of the ratings in the three groups of cities differed. We can offer one speculative explanation of this which requires some further investigation.

The ratings in each city are samples of the proportion of all homes viewing, drawn at various instants in time; the rating refers to an interval made up of many such instants. A rating of 20 per cent, for example, refers to a time interval in which many people tuned in, many tuned out and an average of 20 per cent viewed the show.

Now it would be a great simplification to suppose that a TV audience is made up of only two classes of people: sustained viewers and knob-twiddlers. More accurately we could characterize the TV audience as consisting of people who are retuning at different rates. If all of them were sustained viewers then there would be no variation in the ratings over time; if all of them were active knob-twiddlers the ratings would vary considerably from one moment to the next. We must remember, too, that this rate of knob-twiddling together with the characteristics of the show will determine the show's rating.

In those cities without tune-in advertising the usual full spectrum of tuning habits may be in existence. The "Show of the Month" gets an average rating of about 14 per cent but the actual percentage of viewers at various instants of time oscillates around this figure as people tune in and tune out. The ratings from seven such cities reflect this variance in each separate rating.

With tune-in advertising, however, some of the knob-twiddlers are saved from twiddling; some determine in advance to become sustained viewers, others are repelled beforehand. In short there are now less twiddlers left to twiddle; consequently the ratings will show less moment to moment variation. This in turn reflects itself in less variance between the ratings drawn from those cities subjected to tune-in advertising.

Apparently, advertising brings people into the audience who would like to see your show, but at the same time it screens out others who may not. If an advertiser can pre-judge the popularity and drawing power of the show he is presenting on television, then these results indicate the following. If the show is quite popular and a large proportion of the viewing public should want to see it, then tune-in advertising should increase the total audience. If a show is judged to be unfamiliar to the public and the title has little drawing power, the wisest course of action may be not to advertise and let chance viewing tendencies take over.

## CONCLUSIONS

Tune-in advertising had no effect on the number of sets viewing a low-rated "Show of the Month."

We infer that tune-in advertising does select an audience—both by attracting viewers and by driving them away. It may be a highly effective method of getting an audience who want to see your program, and eliminating people who are "trapped" into seeing it.

We further infer that tune-in advertising may actually be wasteful if you wish to expose your advertising message to individuals who would not normally watch the show.

# FEDERAL STATISTICS IN ADVERTISING

INGRID C. KILDEGAARD
**ARF Research Statistician**

SOUND ADVERTISING STRATEGY requires prediction. What sales do you expect from impulse buyers, from buyers who plan ahead, from those who merely need replacements? How many are predisposed buyers to be directed to your brand, and how many must also be directed to your product group? The answers rarely exist, so you still have to guess.

But if you know how many people own your brand and how many say they intend to buy it, your predictions are bound to improve. Happily, some of this latter information is available from the Bureau of the Census.

### *Consumer Possessions*

A rich vein in the 1960 Census gold mine will be the data on household possessions. The Bureau plans to release 51 advance reports—for 50 states and the District of Columbia—on the number of owning households by state, county and standard metropolitan statistical area. While the information was collected from a sample in the 1960 Census of Housing, the sample size is adequate for providing reliable estimates for counties over 2,000 in population.

Here are the nine possessions to be reported:

Air conditioning (one room unit, two or more room units, central system, none)
Automobile available (one, two, three or more, none)
Clothes dryer (gas heated, electrically heated, none)
Clothes washing machines (wringer or spinner, automatic or semi-automatic, washer-dryer combination, none)
Heating equipment (steam or hot water, warm air furnace, floor wall or pipeless furnace, built-in electric units, other means with flues, other means without flues, none)
Home food freezer (one or more, none)
Radio sets (one, two or more, none)
Telephone available (yes, no)
Television sets (one, two or more, none)

These county estimates will be of obvious value to the advertiser of any of these products (or one closely related, as is packaged bleach to washing machines). Knowing market saturation, he can assign sales potentials, apportion advertising expenditures and plan promotion according to the needs and opportunities of individual areas. The ownership data will help the researcher who wishes to select sets of comparable areas for advertising experiments.

Of more general interest will be the information on radio and television ownership by area. Advertisers who buy or conduct telephone surveys will find the estimates of telephone households of especial interest.

The reports will be issued from May through October 1961. Called *Advance Reports of Selected Equipment and Facilities, 1960 Census of Housing, Series HC (A4)*, they can be ordered from the Superintendent of Documents, U. S. Government Printing Office, Washington 25, D. C.

### *Consumer Buying Intentions*

More consumers said they were going to buy automobiles and television sets within six months in October 1960 than in the previous July. On the other hand, fewer said they planned to buy washing machines and refrigerators. These differences could not be attributed to changes in consumer finances, for reports of past and expected changes in income were comparable for the two months.

These interesting findings are from the Quarterly Survey of Consumer Buying Intentions, conducted by the Bureau of the Census for the Federal Reserve System, and reported in the *Federal Reserve Bulletin*. Begun in January 1959, the survey reports consumer intentions and attitudes toward purchases of durable goods. The first year covered intentions and actual purchases of new and used cars, television sets, refrigerators and washing machines. In 1960 the list was expanded to include houses, air conditioners, clothes dryers, dish washers, and phonograph and radio equipment costing over $100.

In brief, survey results show seasonal patterns in buying plans, with greater variation between quarters than between years. Comparisons of buying plans and purchase rates show differences by income and age groups, though seasonal patterns are apparent within each group. The incidence of purchase among planners is several times greater than among non-planners. But most consumers belong to the non-planning group and their purchases account for the greater proportion of the total. How closely plans and purchases correspond depends upon the item considered, the time interval specified and the planner's degree of certainty.

While it would be naive to translate intentions directly into sales, there may be relationships between sales and intentions which permit the use of these data for predictive purposes. Knowledge of such relationships could aid the advertiser immeasurably.



THE UNIVERSITY OF MICHIGAN LIBRARIES

# ON METHODS

GWYN COLLINS
**ARF Research Mathematician**

This note is prompted by a recent issue of the *Public Opinion Quarterly*[1] which was entirely devoted to "attitude change." It consisted of eight papers on this and related topics contributed by psychologists and sociologists distinguished in this field. The contents of the *Public Opinion Quarterly*, which will be familiar to many readers, are literary rather than analytical in temper, speculative rather than empirical in content. We will consider an alternative approach to this same subject based more on the daily necessity of using *measures* of attitude change than by the desire to organize the *concept* of attitude change into some theoretical structure.

First, and obviously, attitude change implies the measurement of attitudes on at least two occasions. For this reason, we will be mainly preoccupied with the problems of measuring attitudes and giving some meaning to what we measure. There are many definitions of attitude current. These are just a few:

> A neuropsychic state of readiness for mental and physical activity.
>
> The specific mental disposition toward an incoming (or arising) experience, whereby that experience is modified; or, a condition of readiness for a certain type of activity.
>
> A mental and neural state of readiness organized through experience, exerting a directive or dynamic influence upon the individual's response to all objects and situations with which it is related.

In day to day practice, of course, this kind of definition does not help very much. It represents what the psychologist would like to measure or thinks that he measures; not what he does measure. These "neuropsychic states," "mental dispositions" and "neural readinesses" have no physical existence but are inferred from some kind of observation. Since they convey no information that the observation itself does not, we can, without loss, concentrate on the observation itself and eschew psychological inferences about what we think our measure "means." Rather we will apply ourselves to what it will do for us.

Attitude scales are used commercially because they enable us to predict an individual's course of behavior. If a scale is found to be unrelated to the form of behavior we want to predict, we quickly abandon it. By the same token, we choose of several possible attitude scales that which enables us to make the surest prediction.

It may be that a certain form of question, a special kind of interview, or an ingenious kind of interview situation exerts a strong intuitive appeal. We might believe that one or other of these will give us insight into what the respondent "really" feels, "really" wants or is "really" thinking. But the aim of our scale is not to satisfy the intuition but to give high probabilities to predictions. "Depth" interviews, projective techniques and other motivational devices may well be powerful tools but their usefulness must be demonstrated by a close relationship between the information they produce and the behavior in which we are interested. These relationships have to be established (by some kind of correlational technique) whatever the source of the attitude information and whatever kind of behavior is to be predicted. This is why the use of special techniques such as "motivational" research does not enable us to work with any fewer respondents than we would use in more conventional surveys.

In the past measurements of attitudes have been taken with regard to such diverse entities as Negroes, the Democrats, war, and Campbell's Soup and it has not always been quite clear what kind of behavior these scales sought to predict. Some may have been devised with the conscious intention of expanding an already inflated literature about attitude scaling, though in general their authors would indignantly deny that their scales were unrelated to any form of behavior.

Frequently it would be claimed that the scale was not intended to measure behavior at all, but to measure an attitude that was an underlying source of behavior. This last is an unfortunate phrase for

[1] Volume 24, Number 2, Summer 1960.

it suggests that human behavior as we observe it is a second class phenomenon; a poor reflection of underlying psychological factors. It puts the cart before the horse. Behavior, whether the purchase of a brand, the answer to a question, or a rating of a preference is something that we can observe and study. The underlying "motive" is rather an elaborate name for an observed correlation between different aspects of behavior.

In what follows we will assume that the user of an attitude scale is interested in predicting something very specific, usually purchasing behavior. With this in mind we will consider some of the scales that have been or could be used. We will consider two broad kinds of scales. One kind has been derived from theorizing about the underlying variable being measured. Such scales tend to rely on mathematical manipulations and are frequently awkward in construction, tedious in application. They have long been favorites of the psychologists. The other kind just happen to work. The better to appreciate the delights of the latter, we will first spend some time with the complexities of the former.

### *Paired Comparisons*

Probably the most elementary of all scales is generated by asking the respondents to compare several products, two at a time. The procedure is so simple that it is often not regarded as an attitude measure at all but even this method reveals some intricate problems.

Suppose the object of using this procedure is to predict the relative sales of each product. It is reasonable to hypothesize that those who prefer one product to all others will be more likely to purchase it and, in fact, we might expect the franchise of each brand to be equal to the proportion of people preferring it above all other brands. But life being what it is, some brands are not available in some places, some people change their minds and the position becomes a little more complicated. For this reason we become interested in people's second, third and other choices in order to improve our predictions.

Suppose now that we ask the first respondent about only four products A, B, C and D. If he prefers A to B we will denote it by $A > B$. His results might be $A > B$; $A > C$; $A > D$;
$B > C$; $B > D$;
$C > D$.

The judgments of this respondent enable us to put his preferences for the four products in rank order: $A > B > C > D$. We know, or we can usually show, that such a respondent will be more likely to buy A than its competitors, and if he does not buy A, is more likely to buy B than either C or D. Moreover, our degree of interest in his preferences between B, C and D will vary with his likelihood of buying one of these rather than A. In short, we are interested not only that he prefers A to B but also want to know how *much* he prefers A to B.

A difficulty that sometimes occurs in practice with this paired comparisons method is that a respondent's judgments are not consistent. He might, for instance, prefer A to B and B to C yet perversely prefer C to A. One reason could be that in comparing A with B one dimension common to both, say size for example, tends to stand out and his judgment is made on the degree to which they each possess this attribute. The same dimension may also be the most obvious for making judgments between B and C but when A and C are compared the outstanding quality on which they differ may be an entirely different dimension, say color, and this could make for inconsistencies in judgments.

Since the researcher is interested in relating his scale to one aspect of behavior only, say probability of purchase, the existence of more than one dimension of judgment is a nuisance. Often it can be avoided by asking the subject to rank order the products by considering them all at once rather than pair by pair. This, of course, does not remove the problem of having two dimensions mixed up in our scale but it does force each respondent to use only one dimension (though different respondents may use different ones) and often reveals two or more dimensions at the pilot stage of a survey by a respondent's complaints about the difficulties that he has in performing the ranking.

### *Rank Order*

The kind of procedure we have described only succeeds in putting products in rank order. From it we can have no idea whether the choice is a mere toss-up or the difference between night and day. We can give numbers to the rank order scale we have described by giving the product we like best one point, the next two points, and so on. But in assigning such numbers we may forget that though the difference between 3 and 4 is the same as between 1 and 2 the difference between product A and B might seem to the respondent very great compared with the difference between C and D.

This raises an issue of great importance in attitude scaling generally; the assignment of numbers to preferences or degrees of favorableness and unfavorableness. The number system with which we are familiar from elementary arithmetic has several properties which we have long taken for granted. Scales of measurement, particularly attitude scales, do not always have them and arithmetical manipulations with attitude scales can sometimes be misleading. Three special properties of numbers, as we know them from our elementary arithmetic, serve to distinguish different kinds of scales.

### *Three Kinds of Scales*

First, the number system is ordered; any one number is always greater than, equal to, or less than each other number. This is the property which encourages us to attach numbers to things which can be put in rank order. A scale which goes along with the number system this far is usually described as an *ordinal* scale.

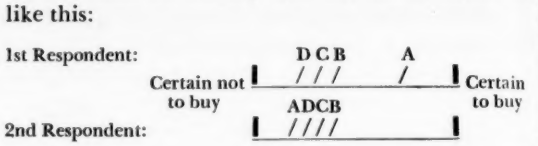
Second, the interval between any two numbers is itself a number. Because these intervals are numbers they must be ordered (by the first property) and this means that the intervals between numbers can be compared with one another. Scales with this second property are often called *interval* scales. Measurements on this kind of scale are like points on an endless string. We can tell which order they are in, we can tell the distance between them and we can even find the average positions of groups of points, but we cannot tell how far any of them are from the beginning of the string.

Then some scales of measurement such as those for weight and length are such that the ratio between any two numbers on the scale is a number. This makes it possible for us to speak of one weight as being twice as heavy as another, of one length as four times as long as some other length. It can be shown that scales of this type, generally known as *ratio* scales, always have a zero point corresponding to the total absence of the quantity being measured.

Returning to our example of rank order scales, we see that they are ordinal but not interval scales. With them we cannot sensibly add the ranks given to products by several respondents. If we try we are misled. Suppose, for example, that we have only two respondents and they rank the four products of our example like this:

| | |
|---|---|
| 1st Respondent | $A > B > C > D$ |
| 2nd Respondent | $B > C > D > A$ |

Now the actual probability of the two respondents buying the various products at the next purchase, which we usually will not know, could be like this:




If we assign scores to the ranks, say 4 to the first rank, 3 to the second, 2 to the third and 1 to the last we would, on adding the points for each product, get:

$$\begin{aligned} &A \ldots\ldots 4+1=5 \\ &B \ldots\ldots 3+4=7 \\ &C \ldots\ldots 2+3=5 \\ &D \ldots\ldots 1+2=3 \end{aligned}$$

This will make it look as though B is more likely to be purchased than A when from our diagram, which contains information we usually do not have, we know that this is not the case.

Several attempts have been made to make rank order or paired comparisons scales into interval scales so that the answers from different respondents can be averaged. To do this we must find some way of ranking the intervals between scores in magnitude. One method is suggested by what is known as Thurstone's Law of Comparative Judgment. It states that in assigning scale positions to two products, A and B, the size of the interval between them will be the product of two factors: the first depends on the frequency which A is preferred to B; the second is a function of the variability of judgments about A, the variability of judgments about B, and of the way in which judgments about A and B are related. Some assumptions are usually made to simplify this Law. First, the variability in judgments about each product is assumed to be the same; second, it is assumed that the common factors, if any, present in making pairs of such judgments always operate to the same extent.

The effect of these assumptions is that the Law of Comparative Judgment reduces to a rule, namely, the distance between the scale positions of two products A and B will depend on the percentage of judgments in which A is preferred to B. The distance between them is actually derived in this way: first, the block of proportions showing how often each product is preferred to each other is written out like this:

| | A | B | C | D |
|---|---|---|---|---|
| A | — | .49 | .22 | .05 |
| B | .51 | — | .38 | .13 |
| C | .78 | .62 | — | .35 |
| D | .95 | .87 | .65 | — |

where the figures in the table show how often the product at the top of the table is preferred to the product at the side of it.

Each of these proportions is then converted into the deviates of the normal distribution corresponding to them. The table then becomes:

| | A | B | C | D |
|---|---|---|---|---|
| A | ... | —.05 | —.79 | —1.65 |
| B | .05 | ... | —.30 | —1.13 |
| C | .79 | .30 | ... | — .39 |
| D | 1.65 | 1.13 | .39 | ... |
| Sum of columns: | 2.49 | 1.38 | —.70 | —3.17 |
| Average of columns: (= scale value) | 0.62 | 0.35 | —.18 | — .79 |

and the scale values are calculated as indicated in the second table. These values merely indicate the distance between scale points and we can, in the above example, add .79 to each to make all of them positive, for we can choose the zero point in a completely arbitrary way.

That the difference between the scale position of two products should depend on the number of times one is preferred to the other seems a reasonable assumption to make. If all comparisons are made on the same dimension then when an overwhelming number of people prefer A to B we can guess that the difference in their market popularity will be great. But when there is only a slight preference for A over B we might feel that there will be little difference in their market popularity. (We might go a stage further and guess that products A and B were so much alike that respondents had a hard time distinguishing between them and chose one rather than the other largely by chance. To be sure about this we would want to separate those who preferred A to B from those who preferred B to A and to consider whether we were dealing with two different kinds of people with different kinds of characteristics.)

This gives us a way of showing preferences for several products or different brands of a product field on a scale. It shows us not only which is preferred but by how much. But it must be emphasized that if this scale does not correlate with actual behavior of some kind, e.g., dollar market share, it will not be useful. Now this is something to be determined by experiment, probably in several product fields. All that can be said for the scale in the absence of such experiments is that intuitively it commends itself as worth trying and as more reasonable than the rough practice of adding ranks together.

### *Equal Appearing Intervals*

The same thoughts which gave rise to the Law of Comparative Judgment have inspired two other kinds of attitude scaling methods, both very popular in the psychological literature: the method of equal appearing intervals and the method of successive intervals. These methods are more often used to describe the chance of individuals exhibiting some kind of behavior than they are to describe the chance of a product attracting it. Both methods involve a two-stage process.

In both, a large number of statements are collected, all of them about the object of the behavior in which we are interested. A number of respondents, usually 50 or more, are asked to sort each statement into one of seven, nine or eleven piles ranging in degree of favorableness to some kind of behavior. The sorters are instructed to make the difference in favorableness between consecutive piles appear equal. In the method of equal appearing intervals the scale value of each statement is the average of the pile numbers in which it is found. The attitude scale itself is made up of a selection of some ten to 20 statements representing a wide range of scale values and each of which was given a fairly small dispersion between different piles. The scale is then ready to take into the field and respondents are given scale scores equal to the average scale values of all the items with which they agree.

The big question, of course, is whether the first set of respondents can really put a series of statements into piles such that assent to them represents equally spaced out probabilities of, say, purchasing action. If they cannot, there is no reason to regard this kind of scale as an interval scale.

The method of successive intervals is similar to that of equal appearing intervals but goes one better. It differs from it only in the way in which the scale scores are computed for each statement. In the method of successive intervals, the scale distances between the items are defined in terms of the way in which they were spread between the piles when sorted by the first set of respondents.

### *Rating Scales*

So far we have considered scales which have been constructed with some kind of theory about an underlying continuum in mind. We now turn to scales which are used simply because they are often found to work. These are much simpler in conception and generally easier to administer.

One such is the Likert scale. To construct it we again collect a large number of statements, all of which seem to be related to the action we want to predict. A first set of respondents is invited to express some shade of agreement or disagreement with each statement. Usually they are offered such categories as: strongly disagree, disagree, don't know, agree and strongly agree. These response categories are then scored, say, −2, −1, 0, 1, 2 and each respondent's score is totalled. Each respondent thus has a score for each item and for all the items together. At this stage any one of several processes is used to select those items whose scores closely relate to the total of all item scores. These items then constitute the scale and it is ready to be taken into the field.

By this procedure scales can be devised which correlate well with behavior, though not all scales produced in this way do. It can happen that the total score on all the items taken together is itself of little predictive value. In this case the items constituting the scale will probably be even less good. Passing over this unpleasant prospect we can admire one of the scale's happier characteristics; that of reliability. By this we refer to the way in which the same respondents tend to generate the same score on different occasions. A scale's ability to correlate with some measure of behavior, which we often describe as its validity, obviously cannot be any better than its ability to agree with itself from one moment to the next.

The Likert scale is only one of many which generates numbers describing how a respondent feels about a product or an issue. Other examples, much more frequently used, ask the respondent about the product more directly.

For example, we might ask the question, "How do you feel about X? Do you approve strongly, approve, don't know, disapprove, or disapprove strongly?" These answers can then be scored +2, +1, 0, −1, −2. Again we might ask, "Give me a number from +5 to −5 to show me how you feel about X. If you like X, give it a plus number and the more you like it the bigger the plus number you should give it. If you dislike X, give it a minus number; the more you dislike it the bigger the minus number you should give it."

These and other procedures generate a number for each respondent, but generally the intention is not to predict the behavior of an individual but to predict the behavior of groups of them. To do this we must somehow combine results derived from many respondents and it is chiefly this need that prompts us to assign numbers to respondents' results. But adding these numbers is misleading, as we have already illustrated in discussing the method of paired comparisons.

Sometimes researchers have preferred to work with verbal categories but this only exchanges one knotty problem for another. Consider, for example, a test of a commercial with "before" and "after" attitude measures of the product. It may provide this sort of result:

| | *Before* | *After* |
|---|---|---|
| Like strongly | 25% | 30% |
| Like | 40 | 30 |
| Don't know | 20 | 20 |
| Dislike | 5 | 10 |
| Dislike strongly | 10 | 10 |
| | 100% | 100% |

Did the commercial do the product any good? Even if the two attitude measures were performed on the same respondents, and if there were no conditioning effects arising from asking the same people the same questions, we should still find ourselves asking whether one respondent improved from "like" to "like strongly" was worth two deteriorating from "don't know" to "dislike strongly."

One way of bringing order out of this chaos is to relate the scale directly to the criterion variable of purchase. The following example illustrates this method. We will suppose that we are testing a commercial for some specific product and are using an eleven-point attitude scale to make the evaluation. For convenience we shall suppose that we have 1,000 respondents.

| *Category* | *Before* | *No. of users* | *% of users per category* | *After* | *Predicted no. of users* |
|---|---|---|---|---|---|
| +5 | 350 | 155 | 44.1% | 458 | 202 |
| +4 | 140 | 54 | 38.5 | 124 | 48 |
| +3 | 45 | 8 | 17.8 | 55 | 10 |
| +2 | 28 | 7 | 25.0 | 18 | 5 |
| +1 | 35 | 4 | 11.4 | 30 | 3 |
| 0 | 210 | 12 | 5.7 | 145 | 8 |
| −1 | 27 | 0 | 0 | 18 | 0 |
| −2 | 16 | 0 | 0 | 14 | 0 |
| −3 | 18 | 1 | 5.6 | 15 | 1 |
| −4 | 37 | 0 | 0 | 19 | 0 |
| −5 | 94 | 2 | 2.1 | 104 | 0 |
| TOTAL | 1,000 | 243 | | 1,000 | 277 |

In this method each of the categories from +5 to −5 is regarded as qualitatively different; they

could equally well be identified by names for at no stage do we add them together. In each category of the scale we find the number of purchasers of the product. We then assume that after the commercial the same proportion of people in each category will purchase the product. Thus in the table above, looking at the +5 category we find that 155, or 44.1 per cent of respondents out of the 350 in this category use the product. After the commercial there were 458 respondents in this category and if we assume that 44.1 per cent of them will use the product this gives us a predicted 202 users. In the tabulated illustration the commercial resulted in a predicted increase of brand share from 24.3 per cent to 27.7 per cent at the times of measurement. The method is obviously the same whether categories are verbal or numerical.

This kind of treatment might work well when the number of users of the product is quite large. Very often, however, the product has a small brand share which makes it impossible to compute a dependable figure for the percentage of users in each category. In such cases the percentage of users per category might be derived from other brands in the same product group. Another difficulty arises when, in an example like ours, many respondents move into the +5 category not only for the brand in which we are interested but also for its competitors. Application of the method as described could lead to the total of predicted brand shares becoming greater than 100 per cent. This can be avoided by using an attitude scale which limits the number of responses that a subject can make in a single category and which forces the respondent to rank order his favored choices. Asking the respondent to allocate a fixed number of points between all brands, and taking the number of points allocated to each as its scale value, appears a good device for this purpose.


***Summary***

Some qualities that have been discussed must appear in any attitude scale that is to be useful. First, it must have some consistency over time. If the scale is applied twice to the same respondent it should generate the same, or almost the same, result. This kind of stability, or reliability, is a necessary but not sufficient quality in any measuring device.

Then the scale should be valid, which is to say that it must relate to some observable behavior. But this is not just one simple quality; it has many aspects. The scale must not, for example, relate to two or more behaviors at a time for otherwise we can never tell which of the two is being measured.

Since the scale is to relate to behavior it is usually desirable for the scale to finely discriminate between groups of people who behave differently or have different probabilities of behaving in some way. This means there must be several scale positions available for respondents though reliability of the scale will tend to fall off as the number of categories is increased. The rule is that the variation in behavior between adjoining positions on a scale should be great compared with the variation of behavior within each category.

Most attitude scales are used more than once on the same subject. If the procedure is too long, or difficult to understand, or repetitive about a brand or product, the respondent's first answers will influence his answers on the second application.

All these points are each aspects of one central quality, one both necessary and sufficient for a useful scale; that of relating to behavior. It is not necessary that the questions used should appear rational, or that there should be an underlying mathematical model. Such considerations merely offer valuable hints in the construction of a measuring device which is *likely* to work. The one final test of an attitude measure is its ability to predict behavior.


*Science is nothing but trained and organized common sense, differing from the latter only as a veteran may from a raw recruit, and its methods differ from those of common sense only as the guardsman's cut and thrust differ from the manner in which a savage wields his club.*

—THOMAS HUXLEY

# RESEARCH IN REVIEW

GWYN COLLINS, THORNTON C. LOCKWOOD and GEORGE M. SHIREY
ARF Staff

## New Way to Spread Money on Media

**A. M. Lee and A. J. Burkart. Some Optimization Problems in Advertising Media Planning. *Operational Research Quarterly*, Vol. 11, No. 3, September 1960, pp. 113-122.**

This stimulating and useful paper may set a new trend in media allocation studies.

Intended for an audience of operations researchers, it is so phrased as to be unattractive to many media and research men. For them, the following account sketches some of the principal ideas.

The authors have a campaign in which their ads appear only once in each magazine selected. They were able, however, to select various amounts of space in any of a number of such publications. First they sought to maximize frequency; then they tried instead to maximize reach.

They assumed that every page of each magazine was equally likely to be read and furthermore, that the probability of anyone noting an ad on any page was equal to the square root of the proportion of the page occupied by it. (This "square root law," recently referred to in a Japanese publication[1] as "the well known square root law," is based on some experimental evidence presented in a 1914 *Psychological Review*.) Then it was assumed that readers of both magazines A and B are no more likely to see the ad in B from having previously seen it in A.

The authors write down an expression for the number of people receiving 0, 1, 2, 3, etc., impressions from any fixed number of magazines. It is an ugly thing which we will take for granted. The number of people receiving no impressions is then written as $I_o$, the number receiving one and only one as $I_1$, and so on. Each of these is obviously a function of the number of magazines employed and the amount of space used in each. Messrs. Lee and Burkart then write down a simple formula for advertising effectiveness.

[1] Jiro Yamanaka. *A Method of Prediction of Readership Scores*. Tokyo: Dentsu Advertising Ltd., 1960.

E (effectiveness) $= w_oI_o + w_1I_1 + w_2I_2 \ldots + w_nI_n$ where the w's can be given any values we care to attach to the receipt of no impressions, one impression, and so on up to n impressions. If we want E to represent reach we simply set $w_o$ equal to zero and all the other w's equal to one. Similarly for frequency we set $w_o$ equal to zero, $w_1$ equal to one, $w_2$ equal to two and, in general, $w_i$ equal to i.

Now maximizing reach, for example, is no problem with an unrestricted budget. The authors list their constraints as: (a) no more than a full page to be used in each magazine; and (b) the total money to be allocated is fixed and it is assumed that ads of less than full page size will have pro rata cost.

The allocation of fixed resources so as to maximize return, subject to a number of simple constraints, is now a fairly everyday procedure of operations research. Electronic computer programs may be bought ready-made for this purpose.

Maximizing average impressions, involving as it does the total audience of the magazines, is a relatively simple problem, for the equations necessary have only as many terms as the number of magazines considered. The solution provided by Lee and Burkart is: for greatest average impression, advertising space should be bought in each publication in inverse proportion to the square of its cost per thousand.

The reach problem is a lot less tractable because it involves unduplicated readership. To represent it for three magazines an equation with seven terms is necessary. For 15 magazines as used in this article the number of terms would be 32,767. This difficulty could have been greatly eased if the authors had had access to J.-M. Agostini's work described on pp. 11-14 of this *Journal*. Instead they make some radical assumptions. These amount to accepting that all the magazines have roughly equal readership, that the readership of each magazine overlaps the readership of each other in purely chance fashion, and that duplications of higher than the third order are fairly insignificant.

The authors carried out each calculation by starting with the allocation for greatest frequency and then moved by successive approximations to the best allocation for reach. Surprisingly these solutions were very close to one another. This prompted the authors to explore a rule to determine the conditions under which the solution of both problems became identical. Their conclusion is that if the magazines' readerships overlap in a chance way, and if the ratio of the square of the readership of each to its cost per thousand is even roughly equal for each of the magazines, then the optimum reach and frequency are obtained by buying space in each publication inversely proportional to the square of its cost per thousand, provided each ad is no more than full page.

The authors are now applying their ideas to a series of ads in one or more publications, decay and acquisition effects, television, combinations of print and TV, and the spacing of ads and commercials over time. —GWYN COLLINS

# Now You See It, Now You Don't

**Edwin B. Parker. Subliminal Stimulation and Voting Behavior. *Journalism Quarterly*, Vol. 37, No. 4, Autumn 1960, pp. 588-590.**

The literature on subliminal stimulation (subliminal perception, subception, learning without awareness, etc.) has devoted a great deal of attention to attitudes and behavior generated from so-called subliminal stimulation, but considerably less to the stimulus itself.

Any stimulus may be defined in terms of its physical properties such as duration, intensity and frequency. However, as ARF's *The Application of Subliminal Perception in Advertising* points out, the limen (threshold) is a statistical concept usually defined as that stimulus level at which perception is claimed by 50 per cent of the group exposed. Therefore a subliminal stimulus must be less than that level, but presumably perceptible by at least some of the group.

Furthermore, perception of a subliminal stimulus varies among different individuals, and with the same individual at different times. Perception of communication also depends upon the individual's psychological predisposition. For example, visual messages transmitted subliminally by television would obviously have little effect, regardless of their duration, intensity or frequency, on those respondents only *listening* to television, i.e., reading a book or out of the room.

Recent reports of subliminal perception have not always described satisfactorily the stimulus in the response terms mentioned above. Reporting a stimulus as subliminal does not make it so. A stimulus may be so subliminal as to be imperceptible to all. Writers on the subject agree that stimulation can be considered subliminal for any given respondent, if the individual is not "consciously aware" of the stimulus, but do not agree as to what constitutes "conscious awareness."

In this article the author's experimental design provided for four groups to be exposed "subliminally." Two were motivated by being told that a film, together with a trait list for each of two candidates, provided the minimal information needed to distinguish between an excellent and a potentially dangerous school board member. All four were told that the film and trait lists were being studied to find out what propaganda effects could be attributed to them. Half the motivated and control groups were exposed to a "Vote G" message, and half to a "Vote H" message. The design apparently precluded any control of the level of subliminality.

The author disposes of this question easily enough: the stimulus, he reports, is a "subliminal message." He gives its physical dimensions: bulb and projector wattage, shutter speed and aperture. The author states that the "stimulus intensity permitted the message to be seen occasionally by the experimenters. None of the subjects reported seeing anything unusual during the experiment." How they reported this is not specified. He summarizes:

> These results, coupled with the negative results of other researchers, lead to the conclusion that subliminal stimulation is at best an inefficient means of communication and that techniques of subliminal stimulation are unsuited for the purposes of mass persuasion. This conclusion must remain tentative, however, since it is always conceivable that someone will be able to find some conditions under which it will work effectively. But until such conditions are found we have a straightforward reason, without raising any ethical questions, for not using subliminal persuasive techniques: They don't work.

Was the stimulus too subliminal to produce an effect? Would longer or more intense exposure have affected voting behavior? While not disagreeing with his results, some readers may think the author's conclusions premature.

What kind of evidence is needed to verify a subliminally stimulated effect? Assume an experimental group of respondents exposed subliminally (operationally defined as above) to advertising, and a control group not exposed. If the two groups do

not differ significantly with regard to claimed awareness of the stimulus, and if they *do* differ in the expected direction with regard to attitude change (using appropriate measures of claimed awareness and attitude), then evidence for successful subliminal persuasion exists. Congress and advertisers would probably be equally intrigued by such evidence. As far as we know it does not exist.

—THORNTON C. LOCKWOOD

## Attitudes Are Answers to Questions

**Hugh Hardy. Study Shows Public Attitude Toward Advertising is Surprisingly Good. *Advertising Age*, Vol. 31, No. 49, December 5, 1960, pp. 77-82.**

Last July the Politz bi-monthly Omnibus attempted to measure public attitudes toward advertising. This report of that study can be conveniently divided into two parts.

The first reports answers to: "In general, which one of these statements best expresses your own attitude toward advertising?" The author "condensed the seven-point scale of attitudes to three on the charts for simplicity," and found 17 per cent "generally unfavorable." The results of other questions, which the author says pertain to various aspects of believability, are then taken up. Negative attitudes toward advertising expressed in such terms as false, deceptive, misleading, untruthful, irritating, distasteful and unfavorable are ravelled and unravelled. This all leads to the lengthier second part of the study which is well summarized by the author as follows, ". . . if a more definitive study were to be made, it should make an effort to separate questions of taste, irritability, and believability from the question of deception."

Mr. Hardy labels most antagonists of advertising as "intellectuals." For convenience, he lumps into this category anyone who expresses negative attitudes toward advertising. Some readers may sympathize with his motives but disagree with his definition.

The author's main conclusion is built around the 17 per cent reported to be "generally unfavorable." But some readers may ponder the author's contention that the study "was geared to negative responses . . . intentionally [to obtain] findings [with] a virtual maximum of criticism." Why this approach was desired is unclear, since the author himself says that "nowhere did we see any surveys that could be considered a reliable and objective expression of public likes and dislikes." Whether this objective was fully implemented is another matter; the only question reported (see above) does not appear to be deliberately constructed to produce a "maximum" of negative answers. Perhaps something else in the questionnaire produced the desired effect.

Those who carefully read this article will realize that measuring attitudes toward advertising is an extremely complex business. To them Mr. Hardy has done a real service. It is unfortunate that a more appropriate title and introduction weren't used to put across this most important result of the survey.

For attitude measures are rarely useful by themselves. The difficulty of using them increases when the relationship between attitudes and behavior is ignored. Mr. Hardy's study has shown some differences between the attitudes toward advertising of various subgroups in the sample. But is advertising also less *effective* with men, post-graduates and people 65 years of age and older? Are people who claim to have seen false or misleading advertising less influenced by advertising? Do respondents who recall the product class and specific brand involved in their claim of false or misleading advertising never buy these products and brands? Clearly the advertiser could use answers to such questions. With them he probably would be more confident of handling the political and social issues advertising involves.

In the absence of behavioral validation, measures of public attitudes toward advertising may provide some value by being asked repeatedly over time. This would at least reveal trends in answers to the questions asked.

The author is to be commended for tackling the attitudes-toward-advertising tiger. At the conclusion, however, the issue of who has swallowed whom is still in doubt. —THORNTON C. LOCKWOOD

## If at First You Don't Succeed, Wait

**Schwerin Research Corporation. When Should the Effects of Television Advertising be Measured? Part 2: Changes in Attitude and Behavior. *Technical and Analytical Review*, No. 5, Summer 1960, pp. 1-14.**

In advertising research, attitudes are frequently distinguished from behavior by the techniques used to measure each. An attitude is an indirect and subjective measure in that it requires a verbal or written report by the individual. Behavior may be

measured directly and objectively, with limited, if any, reporting by the individual. If the reader applies these definitions to the title of this pamphlet, he will find no evidence of behavior changes resulting from television advertising.

What we have instead is a description of seven studies, each of which provides evidence of attitude changes resulting from various kinds of communications. Four of the projects concern political communications quite remote from advertising, and three concern television commercials. The former are drawn from professional journals and present evidence of the persistence of attitude change, the sleeper (delayed) effect resulting from low communicator credibility, the effect of reinstatement of the communicator on the sleeper effect, and the reinforcement of attitude change resulting from repetition.

This is a well written beginning and interestingly presented. The author is ready to take up television advertising. Perhaps his warning that "these data are not structured either to confirm or deny the hypotheses suggested by the academic data" adequately prepares some readers for the subsequent let-down.

Interpreting Schwerin Research Corporation data, the *Review* states that a TV commercial "wears out" when it motivates a smaller share of the audience to want the advertised brand, as measured by the competitive preference technique. Competitive preference increments are reported for 31 anonymous products after various periods of advertising. This brief consideration of wearing out suggests that the advertiser is interested only in the increment, and not the level, of brand preference at different points in time. Many advertisers would probably have been interested in finding out whether levels as well as increments of brand preference differed significantly.

The author next discusses the sleeper effect of television advertising, and presents data which suggest that seasonal influences and intervening campaigns may cause this effect on attitude change. The key word here is "suggest." Inadequate control of frequency of exposure to commercials, the probable noncomparability of samples, and the variation in time intervals preclude any general conclusions.

One value of this pamphlet should not be overlooked. In advertising research, where security precautions often seem unnecessarily excessive, research proposal writers suffer from the lack of exposure to the thoughts and experiences of other practitioners. With the excellent background provided by this pamphlet, some readers may be stimulated to answer the author's plea for "structured research [in order to examine thoroughly] the occurrence of a delayed increase in commercial effectiveness."

But consolidating findings, however laudable in intent, is not always useful. It is difficult to interpret differing results, obtained by using different measuring devices and administering different messages to people with different predispositions. The author proposes an uneasy marriage between academic and commercial research efforts as a way of determining study objectives. The reviewer sincerely hopes research therapy will save this marriage.

—THORNTON C. LOCKWOOD

# Necessary But Insufficient

**Joseph T. Klapper. *The Effects of Mass Communication.* Glencoe, Illinois: The Free Press, 1960. $5.00.**

Mr. Klapper has collected published findings and provocative conjecture concerning the social and psychological effects of mass communication. He has reviewed some 1,000 studies, essays and reports and selected over 270 to include in his book.

They are organized around five generalizations:

1. Mass communication *ordinarily* does not serve as a necessary and sufficient cause of audience effects, but rather functions among and through a nexus of mediating factors and influences.
2. These mediating factors are such that they typically render mass communication a contributory agent, but not the sole cause, in a process of reinforcing the existing conditions.
3. On such occasions as mass communication does function in the service of change, one of two conditions is likely to exist. Either: (a) the mediating factors will be found to be inoperative and the effect of media will be found to be direct; or (b) the mediating factors, which normally favor reinforcement, will be found to be themselves impelling toward change.
4. There are certain residual situations in which mass communication seems to produce direct effects, or directly and of itself to serve certain psychophysical functions.
5. The efficacy of mass communication is affected by various aspects of the media and communications themselves or of the communication situation.

The first part of the book is concerned with mass communication as an agent of persuasion, investigating its effect as a reinforcement agent, its efficiency in creating opinions and attitudes on new issues, and its ability as a conversion agent. The second half of the book deals with research on the effects of certain specific media content, e.g., media

treatment of crime and violence, the effects of "escapist" material, the effects of adult television programs on child audiences, and finally, the alleged dangers of the passive occupation of media attendance. After each chapter the author summarizes the findings reported.

The problem of organizing and summarizing 270 studies and of making the finished product readable is a difficult one which has been successfully solved by Mr. Klapper. He includes the work carried out in Great Britain on the effects of television, and he elaborates on several areas of communication research he only briefly touched on in an earlier work of the same title.

In reading the book, it becomes obvious that the author is well acquainted with the studies. From time to time, but all too rarely, he draws on his experience and injects his own thoughts on some aspect of the subject.

For the most part, Mr. Klapper draws heavily on the views of others. This reviewer regrets that we did not have more of Mr. Klapper, for his insights and interpretations are among the most stimulating passages of the book. This reviewer hopes he will find it possible to contribute a further work which will emphasize his own theories. As it is, this book will be an asset for all who are interested in mass communication. —GEORGE M. SHIREY

# Something Old, Something New

**Walter Gruen. Preference for New Products and Its Relationship to Different Measures of Conformity. *Journal of Applied Psychology*, Vol. 44, No. 6, December 1960, pp. 361-364.**

The middle class American seeks the approval of his peers by keeping up-to-date. So say Whyte, Fromm and Riesman, and the author of this study fears that as a result the advertiser may stress "too frequently the newness of his product as a sales appeal." He is not sure that the encouragement of such "subjective obsolescence" is entirely justified.

The author translates his concern into a specific hypothesis: "some people prefer new styles and new exteriors regardless of the type of product." His experimental design further translates "some people" as 114 college psychology students, "newness" as two ranges of years, 1950-56 vs. 1957-59, and "product" as foreign cars, silverware patterns, dinnerware, spring suit ensembles, dresses, living room furniture, office furniture, project house models, and labels from canned soup and instant coffee.

Three fairly evenly matched pictures were selected to represent each product and made into slides. Two were of the same brand with one carrying an old label (a year between 1950-56) and one a new label (a year between 1957-59). These intervals were varied purposefully from three to seven years to help disguise the study's purpose (presumably from the respondent). For each pair of labelled slides, the years were reversed for half the respondents. The third slide was of a different brand with no year label. Appropriate refreshments were served during the presentation of soup and coffee slides to provide grounds for judging these two items. Respondents were carefully instructed to "judge each object on the slide, or each food sample according to how you *feel* about it *now* . . . not what you may know or think about the object."

Rather than test the hypothesis directly by asking which picture was preferred, the author used the semantic differential scales (i.e., good-bad, beautiful-ugly, clean-dirty and pleasant-unpleasant) Osgood found to be most closely associated with his evaluative factor. Although the author states that the semantic differential was used to obtain a more sensitive measure of the reaction than is possible by asking for preference, 58 per cent of the sample showed no particular preference for either the "old" or the "new" products.

Evaluation scores of each respondent's predisposition toward newness were determined for each of the ten product categories. In determining these scores and analyzing his results, the author weighted equally the individual polar adjective scales and the ten individual products, and applied the same ten-year scale of newness to all products. Many readers will believe that some justification for this weighting should have been provided.

Four additional tests of conformity were administered to the subject: Gross's other- vs. inner-directed scale, Moeller and Applezweig's social approval or self-approval scale, agreement-disagreement on an American core culture inventory, and Asch's scale of agreement or disagreement with an electrical stooge's decision. The author found no significant relationship among his measured preference for newness and these four tests.

In summary, the author seems to have taken Whyte, Fromm and Riesman too literally, too seriously, or both. This reviewer, for one, is hard put to detect subjective obsolescence in, say, silverware patterns from 1950 to 1959.

—THORNTON C. LOCKWOOD

# PUBLICATIONS RECEIVED

**MILDRED CANTRELL**
**ARF Librarian**

The ARF Library is proving to be a valuable information center. It contains an outstanding and in many ways exclusive collection of reports and clippings on research concerned with advertising, marketing, media, copy, motivation and sales effectiveness. Since its establishment in 1953, the collection has grown rapidly. With this growth has come a tremendous increase in usage by Foundation members. Non-members, who are becoming more aware of our highly specialized and extensive collection of materials on advertising research and related topics, are turning to us for this often difficult-to-find information. These requests are filled whenever possible, but preference is given to members in all cases.

If you wish to use the library, please call to find out if the material you need is available. If it is, an appointment to visit can be made.

Books, research reports or ARF publications are also available through inter-library loans. These loans should be requested in writing.

Acquisitions since the last issue of this *Journal* (December 1960) are listed below. Permanent copies can be obtained directly from the publisher; loan copies are available to our members.

Anns, John, *Who's Who in Market Research*. Lausanne, Switzerland: John Anns Publishing Company, 1960. 244 pp. $10.00.

Brady, Dorothy and Eleanor M. Snyder, *Relation of Family Status to Income in the Labor Force, New York State, 1956*. New York: New York State Department of Labor, January 1960. 19 pp. Free.

Britt, Steuart Henderson, *The Spenders*. New York: McGraw-Hill Book Company, Inc., 1960. 293 pp. $4.95.

Bureau of Advertising, *Newspapers at Your Service*. New York: Bureau of Advertising, 1960-61. 35 pp. Free.

Dahl, Robert, Mason Haire and Paul Lazarsfeld, *Social Science Research on Business*. New York: Columbia University Press, 1959. 185 pp. $3.00.

Deutschmann, Paul J., *Newspaper Content of Twelve Metropolitan Dailies*. Cincinnati: Scripps-Howard Research, September 1959. 156 pp. $4.00.

Fetuiler, Leon, *A Theory of Cognitive Dissonance*. Evanston, Illinois: Row, Peterson and Company, 1957. 291 pp. $5.25.

Hancock, Robert S. (Editor), *Dynamic Marketing for a Changing World*. Chicago: American Marketing Association, 1960. 570 pp. $5.00 to AMA members, $7.50 to non-members.

Judkins, Jay, *Directory of National Associations of Businessmen, 1960*. Washington, D. C.: U. S. Government Printing Office, November 1960. 81 pp. $.50.

Klapper, Joseph T., *The Effects of Mass Communication*. Glencoe, Illinois: The Free Press, 1960. 302 pp. $5.00.

Klein, B. and Company, *Guide to American Directories*. New York: B. Klein and Company, 1960. 380 pp. $20.00.

Levenson, J. H. (Editor), *Electronic Business Machines*. New York: Philosophical Library, Inc. 272 pp. $15.00.

Luce, R. Duncan, *Developments in Mathematical Psychology*. Glencoe, Illinois: The Free Press, 1960. 315 pp. $7.50.

Mosteller, Frederick, Robert E. K. Rourke and George B. Thomas, Jr., *Probability and Statistics*. Reading, Massachusetts: Addison Wesley Publishing Company, Inc., 1961. 395 pp. $4.00.

National Analysts, Inc., *Follow the Leader Fact Book*. Philadelphia: The Philadelphia Bulletin, 1960. 43 pp. Free.

National Analysts, Inc., *A Study of Newspaper Readership in Telephone Households in Greater Philadelphia*. Philadelphia: The Evening and Sunday Bulletin, April 1960. Volume 1, 25 pp. Volume 2, 50 pp. Free.

Newcomb, Robert and Marg Sammons, *Employee Communications in Action*. New York: Harper and Brothers Publishers, 1961. 337 pp. $5.75.

Nielsen, A. C., Company, *Report on Los Angeles Pilot Test of Outdoor Advertising Coverage*. New York: Foster and Kleiser, Division of Metropolitan Broadcasting Corporation, October 1960. 63 pp. Free.

Politz, Alfred, Research Inc., *National Automobile and Tire Survey, 1960*. New York: Cowles Magazine, Inc., 1960. 98 pp. $1.00.

Richmond Times Dispatch, *Newspaper Market Data*. Richmond, Virginia: Richmond Times Dispatch, 1960. 75 pp. Free.

Samstag, Nicholas, *Persuasion for Profit*. Norman, Oklahoma: University of Oklahoma Press, 1958. 208 pp. $3.75.

Scripps-Howard Research, *The Climate of Attitude in Cincinnati, Ohio, June-July, 1960*. Cincinnati: The Stephen H. Wilder Foundation, 1960. 240 pp. $15.00.

Solomon, Herbert, *Mathematical Thinking in the Measurement of Behavior*. Glencoe, Illinois: The Free Press, 1960. 315 pp. $7.50.

United States Department of Agriculture, *Marketing Service Programs Conducted by State Departments of Agriculture*. Washington, D. C.: U. S. Government Printing Office, July 1960. Free.

# Agostini's Constant

To laymen the world of science must seem populated mainly by things. Microscopes and telescopes, cyclotrons and Univacs, bathyspheres and bubble chambers—these are the heroes of popular scientific writing. The scientist rarely upstages the machines he uses or invents. When he does, it may be only as the user or inventor, one supernumerary member of a robot cast of thousands.

Many a scientist is probably reconciled to this state of affairs. He may comfort himself with the thought that among his colleagues, at least, there exists a time-honored distinction between technology and science, with the latter a rather more prestigious activity. From his standpoint, the machines are merely means to an end.

But this poses an interesting problem to the scientist human enough to want public acclaim. How can he become known to his fellow citizens? If he invents a gadget, any resulting fame is bought at the price of perpetuating the popular notion of science as subordinate to technology. He would prefer to be associated with his goal, not his tools.

This goal is nothing less than the extension of human knowledge. The difficulty is that such knowledge seldom comes neatly packaged. Instead it is accumulated in darts and rushes, a hunch here, a fact there. Once in a great while he may be able to report a reliable empirical law, a tested and proven theory, a parameter stable enough to be called a constant. All too often it goes unnamed.

We wonder if, when Archimedes leapt from his bath crying "Eureka!" he knew he had found not only the law of specific gravity but the scientist's path to fame as well. Since his day, few others have enjoyed the exquisite thrill of seeing their names similarly immortalized: Mendel's Law, Einstein's Theory, Planck's Constant—there are not many.

If such fame is rare in the physical and biological sciences, it is virtually nonexistent in the infant science of marketing. In our experience, Reilly's Law of Retail Gravitation is the only case, and Reilly himself has modestly disclaimed the scope and precision implied by the analogy. We would prefer to attribute this lack to similar modesty by other discoverers rather than to the absence of findings of permanent value.

So it is with personal pleasure and vicarious pride that we suggest a namer of knowledge from advertising research. He is Jean-Michel Agostini and he reports his finding on pages 11-14. Briefly, he has discovered that the unduplicated audience of three or more magazines can be accurately estimated from the total audience of those magazines and their duplications in pairs. He represents this relationship by a simple equation containing only one parameter; this parameter equals 1.125 for several combinations of both French and U. S. magazines.

Now it is much too early to regard this finding as any more general than M. Agostini carefully reports it. Even in the physical sciences, constants have a way of changing their values as their fields of applicability are extended. But it is perhaps not too early to draw some morals from this finding.

Most obvious is that mathematical models save time and money. If one obtains simultaneously the total and pair-wise duplicated audiences of a large number of publications, then tabulations of the thousands of higher-order combinations may not need to be done. A few combinations will suffice to determine the generality of Agostini's Constant. If it holds, work is eliminated. If it does not, we will still have a sensitive dependent variable on which to observe the possible influence of other factors.

Second, to name this finding will remind us of the proper roles in science of knowledge, men and things. Knowledge is the goal, and when found it should be named as readily as an inventor christens his brainchild, or an advertiser his new brand. A man had the wit to find it, so his should be its name. Machines were merely the means, and shall remain nameless.

Third, and perhaps most vital in the long run, a finding called constant will stimulate further research. The very word both gratifies and challenges the scientist. He knows that an hypothesis has been stated according to the only criterion of a good hypothesis: that it be testable. The lists are open and the gauntlet is down.

—CHARLES K. RAMOND